Macroeconomic Implications of Specialization and Inequality



Lukas Boehnert

Hertford College

Department of Economics
University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy in Economics

Hilary Term 2026

Abstract

This DPhil thesis is a collection of three theoretical and empirical papers studying the macroeconomic implications of specialization and inequality. The first chapter documents that highly specialized U.S. regions have higher per capita income in the short run but persistently lower growth in the long run. I rationalize this trade-off in a dynamic multi-industry model and derive the optimal regional specialization from a constrained-efficient planner's perspective. The second chapter examines the role of income inequality for the transmission of monetary policy among Euro Area countries. We show that countries with greater inequality exhibit weaker responses to monetary policy shocks and have larger service expenditure share. We rationalize this pattern in a New Keynesian model with non-homothetic preferences and heterogeneous sectoral incomes. The third paper investigates macroeconomic consequences of a decline in resource extraction for fossil fuel exporting countries. Constructing a novel dataset for 13 energy-transition minerals and 122 countries since 1950 we document persistent negative effects of large, exogenous extraction declines on real GDP.

Acknowledgements

My PhD has been an incredible and unexpected journey of many ups and downs and personal growth. Above all, it has given me the opportunity to meet amazing people, from whom I have learned continuously and many of whom I now call friends.

First, I am grateful to my supervisor, Federica Romei, for her continuous academic and personal support during my PhD. I thank her for encouraging me to explore my own research interests, and to work hard while always keeping a balance in life.

I also thank Andrea Chiavari, Sergio de Ferra, Andrea Ferrero, Todd Schoellman and Petr Sedláček for their generous support and feedback. My work has greatly benefited from their guidance, and I have learned immensely from interacting with such outstanding scholars.

I have also been fortunate to have had inspiring teachers and mentors throughout my life. I thank Doug Gollin for his invaluable advice. I am also grateful to my piano and music teacher, Altfried Beele. Your lessons have shaped me in so many ways, within and beyond music.

Besides academia, I would like to thank a number of people for their friendship and support. To my fellow PhDs, Giulio Gottardo, Alex Haas, Matt McKernan, Amelie Mennerich, Manos Perdikakis, Tatjana Schulze, and Michael Wulfsohn for countless discussions on economics, research, and beyond. To Eoin O'Sullivan and Jack Mazzotti for making me feel at home in Oxford. To my New Yorkers, Pepa Majkic, Felix Paffhausen, Luise Quaritsch and Leon Tiedemann-Friedl for all the fun and life lessons along the way. To Moritz Dillmann and Leonard Tusch for inspiring me and reminding me that anything is possible. To Malte Milewski, for the pure joy of lifelong friendship. To my high-school "Unbreakable" group of friends, and to my boys from back home, Moritz, Philipp, Timo and Tobias for being a constant anchor.

Finally, and most importantly, I am deeply grateful to my family, especially to my parents, for their unwavering support and love and for making me the person I am.

Contents

Abstract i

Acknowledgements ii

List of Figures v

List of Tables viii

Introduction 1

1 The Regional Specialization Trade-off 3
- 1.1 Introduction . . . 4
- 1.2 Empirical Analysis . . . 8
- 1.3 A Model of Regional Specialization . . . 13
  - 1.3.1 Environment . . . 14
  - 1.3.2 Unregulated Decentralized Competitive Equilibrium . . . 16
  - 1.3.3 The Role of Specialization . . . 17
  - 1.3.4 Constrained-efficient Planner Equilibrium . . . 19
- 1.4 Quantitative Analysis . . . 23
  - 1.4.1 Calibration . . . 23
  - 1.4.2 The Specialization Trade-off . . . 26
  - 1.4.3 Specialization Dynamics . . . 27
  - 1.4.4 Capital Allocation, Borrowing Decisions and Amplification . . . 28
- 1.5 Constrained-efficient Allocation and Welfare . . . 30
- 1.6 Conclusion . . . 34
- 1.7 Appendix - Empirical Analysis . . . 36
- 1.8 Appendix - Model . . . 51

2 Monetary Policy in Currency Unions with Unequal Countries 54
- 2.1 Introduction . . . 56
- 2.2 Empirical Analysis . . . 58
  - 2.2.1 Data . . . 58
  - 2.2.2 Empirical Regularities . . . 59

2.3 Model . . . 63
2.3.1 Households . . . 63
2.3.2 Firms, Production, and Nominal Rigidities . . . 65
2.3.3 Monetary Policy . . . 67
2.3.4 Market Clearing and Equilibrium . . . 67
2.4 Results: Revealing the Mechanism . . . 68
2.4.1 Non-Homotheticity, Inequality, and Expenditure Shares . . . 69
2.4.2 Monetary Policy Transmission: The Role of Income Inequality and Expenditure Shares . . . 71
2.4.3 Model Robustness and Empirical Grounding . . . 76
2.5 Conclusion . . . 82
2.6 Appendix - Empirical Analysis . . . 83
2.7 Appendix - Model . . . 94

3 Economic Consequences of Large Extraction Declines: Lessons for the Green Transition 99
3.1 Introduction . . . 101
3.2 Identification Strategy: Episodes of Exogenous Extraction Declines . . . 104
3.2.1 The Dataset . . . 105
3.2.2 Large, Persistent, and Macro-relevant Extraction Declines . . . 105
3.2.3 Exogenous Extraction Declines . . . 106
3.2.4 Anticipation . . . 108
3.3 Extraction Decline Episodes: A Preview . . . 109
3.3.1 Extractive Activity . . . 109
3.3.2 Economic Activity . . . 110
3.4 Theoretical Framework . . . 111
3.4.1 Model Setup . . . 111
3.4.2 Model Results . . . 113
3.5 Empirical Framework . . . 115
3.6 The Impact of Extraction Declines on Aggregate Output . . . 116
3.6.1 The Typical Dynamics of Extraction Declines . . . 116
3.6.2 Baseline Results . . . 117
3.6.3 Effects Across Industries . . . 119
3.6.4 Heterogeneity Across Countries . . . 120
3.7 The Role of Institutions . . . 123
3.8 Robustness . . . 125
3.8.1 Sub-Sample Analysis . . . 125
3.8.2 Anticipation Effects . . . 126
3.8.3 Long-Term Results . . . 126
3.8.4 Synthetic Counterfactual Method . . . 127
3.9 Conclusion . . . 128

3.10 Appendix - Empirical Analysis 130

References **140**

List of Figures

1.1 The Specialization Trade-off: Historical Example 4
1.2 The Specialization Trade-off after Controls 9
1.3 Non-parametric estimate: Specialization vs. per capita income 12
1.4 Persistence at the region-industry level 13
1.5 Illustrative interaction between specialization and the financial friction . 17
1.6 The specialization trade-off: Data vs Model 26
1.7 The role of financial amplification for growth 27
1.8 Impulse Responses to a transitory industry-specific productivity shock . 28
1.9 Policy functions for a negative one-standard deviation industry-specific shock 29
1.10 Specialization and the tightness of the collateral constraint 30
1.11 Competitive vs constrained-efficient specialization trade-off 31
1.12 Competitive vs constrained-efficient specialization in 1950 32
1.13 Welfare gains for a negative one-standard deviation industry-specific shock 33
1.14 Welfare under the constrained-efficient specialization in 1950 34
1.15 Commuting Zones by 1950 exposure to structural change 42
1.16 Industrial specialization, relative income and long-run growth 43
1.17 Commuting Zones by 1950 specialization 44
1.18 Commuting Zones by 2020 specialization 44
1.19 Dispersion of income p.c. across labor markets 45
1.20 OLS estimate: Specialization vs. per capita income 46
1.21 Persistence at the regional level 48

2.1 Services expenditure shares by household income level. 60
2.2 Income elasticity of consumption shares by COICOP category. 60
2.3 Inequality and Services Expenditure Shares. 61
2.4 Impulse response of real GDP to a contractionary monetary policy shock. 62
2.5 Cumulative impulse response of trade balance (% of GDP) to a contractionary monetary policy shock. 63
2.6 Household-level service expenditure share $\omega_{j,k}^S$ as a function of household expenditure. 70

2.7 Average service expenditure share $\bar{\omega}_k^S$ as a function of income inequality. 72
2.8 Contractionary monetary policy in union of identical countries. Effect on total output, as function of the stationary-allocation service share of expenditure. 73
2.9 Contractionary monetary policy in union of identical countries. Effects of monetary policy shock on consumption of goods, aggregate and of Ricardian household (left panel) and on relative expenditure of Ricardian household (right panel). 75
2.10 Contractionary monetary policy in union of identical countries. Effects on the price of tradable good (left panel) and on service output (right panel). 76
2.11 Contractionary monetary policy in union of heterogeneous countries. Effects of monetary policy shock on total output and trade balance. 77
2.12 Contractionary monetary policy in union of heterogeneous countries. Alternative values for elasticity of substitution. 78
2.13 Contractionary monetary policy in union of heterogeneous countries. Higher degree of non-homotheticity. 79
2.14 Contractionary monetary policy in union of heterogeneous countries. Inequality stemming from goods sector. 80
2.15 Contractionary monetary policy in union of heterogeneous countries. Inequality in both sectors. 81
2.16 Inequality and service expenditure shares across countries. 84
2.17 Service income shares by decile. 85
2.18 Impulse response functions of real GDP to a contractionary monetary policy shock of one standard deviation for high and low service consumption shares controlling for income inequality. 88
2.19 Impulse response functions of real GDP to a contractionary monetary policy shock of one standard deviation for high and low service consumption shares controlling for the share of HtM agents. 89
2.20 Impulse response functions 89
2.21 Impulse response functions of nominal government expenditure to a contractionary monetary policy shock of one standard deviation for high and low service consumption shares. 90
2.22 Non-durables vs durables 90
2.23 Services vs housing expenditure shares. 91
2.24 Impulse response functions of goods and services output to a contractionary monetary policy shock of one standard deviation. 91
2.25 Impulse response functions of real consumption by sector to a contractionary monetary policy shock of one standard deviation. 92
2.26 Cumulative impulse response functions 93

2.27 Impulse response of log goods, and service price indices. 93
2.28 Impulse response functions for individual countries. 94

3.1 Distribution of extraction decline episodes over time. 108
3.2 Exogenous declines in fossil fuel and extractive activity. 110
3.3 Real GDP in levels and in growth rates during the 35 episodes of extraction declines. 111
3.4 Effects of a fully anticipated permanent output decline compared to the effects of a sequence of transitory negative shocks to mining output. . . . 115
3.5 The accumulated extraction decline of an initial 10% drop in production growth accumulates to a 40% decline after 10 years. 117
3.6 Impulse response function of real GDP to a shock episode. 117
3.7 Impulse response functions of domestic sector variables to a shock episode. 118
3.8 Impulse response functions of external sector variables to a shock episode. 119
3.9 Impulse response functions of real value added (in constant 2015 local currency units) by sector to a shock episode. 120
3.10 Impulse response function of real GDP interacted with the size of the resource sector to a shock episode. 121
3.11 Impulse response function of real GDP interacted with per capita income and sector size to a shock episode. 122
3.12 Impulse response function of real GDP interacted with the extractive sector and the manufacturing sector share of value added to a shock episode. 123
3.13 Impulse response function with the institutional quality indicator (-10 to 10) to a shock episode. 124
3.14 Impulse response functions of sub-sample Real GDP to a shock episode. . 125
3.15 Impulse response functions of GDP interacted with a dummy for anticipation to a shock episode. 126
3.16 Average impulse response of real GDP over 20 years to a shock episode. . 127
3.17 Actual versus synthetic counterfactual medians over time. 128
3.18 Synthetic counterfactual for Gabon. 133
3.19 Estimated persistence (auto-correlation) of the binary shock dummy. . . 134
3.20 Impulse response functions of the employment rate to a shock episode. . 135
3.21 Impulse response function of real GDP to a shock episode with linearity restriction. 135
3.22 Impulse response functions excluding one decline episode a time. 136
3.23 Impulse response functions of real value added by sector to a shock episode. 139

List of Tables

1.1 U.S. Regional industrial specialization over time 11
1.2 Model Calibration 25
1.3 Employment shares 36
1.4 Regression table of regression (1.5) 38
1.5 Regression table of regression (1.5) across different industries 39
1.6 Regression table of regression (1.5) with HHI 40
1.7 Regression table: Specialization on growth at different horizons 41
1.8 Regression table: Specialization change on income change 47
1.9 RCA persistence 48
1.10 Agglomeration parameter by industry 49
1.11 Adjustment cost parameter by industry 50

2.1 Coefficient Estimates from Regression (2.1). 61
2.2 Summary statistics on countries income and consumption. 83
2.3 Regression table of equation 2.1. 84
2.4 Classification of COICOP categories as goods or services. 86
2.5 Classification of NACE categories as Goods, Services or Housing 87

3.1 Summary statistics of the exogenous declines in extractive activity by commodity. 107
3.2 Summary of episodes of declining extractive activity by commodity (in averages). 130
3.3 Summary of the 35 episodes of exogenous extraction declines. 131

Introduction

Economic fortunes vary dramatically across regions over time. Over the long run, once-thriving regions may experience profound decline, while other areas rise to prominence. In the short run, identical economic shocks can generate substantial losses in some regions while leaving others largely unaffected. This thesis seeks to deepen our understanding of why regions and countries follow divergent trajectories, both in terms of long-run growth and short-run responses to shocks. In particular, it focuses on the macroeconomic implications of specialization and inequality in shaping these divergent regional outcomes.

Each of the three chapters consists of an independent research paper. In the first chapter, titled "The Regional Specialization Trade-off", I provide new evidence that a crucial factor in explaining long-run regional variation is the degree of *regional specialization* — the concentration of economic activity across industries. Using U.S. Census microdata, I document that regions that were more specialized in 1950 had higher per capita incomes than more diversified regions. Since then, however, these specialized regions have exhibited persistently lower long-run growth. To rationalize this novel empirical fact, I then develop a dynamic multi-industry model featuring two opposing forces. On the one hand, specialization increases productivity and income through industry-specific agglomeration forces that capture the short-run benefits of specialization. On the other hand, both real factor adjustment costs and a financial friction generate long-run costs of specialization by making the reallocation of factors an inherently slow and costly process, especially for highly specialized regions. A calibrated model disciplined by U.S. Census data on all 722 commuting zones is able to capture 51% of the relationship between relative specialization in 1950 and 1950-2020 income growth across the U.S. Finally, I derive the optimal specialization implemented by a constrained-efficient, regional planner that balances inefficiencies stemming from the agglomeration and the financial friction. The constrained-efficient allocation reveals that less specialization can raise welfare by reducing a region's exposure to industry-specific downturns.

In the second chapter, titled "Monetary Policy in Currency Unions with Unequal Countries" and co-authored with Kurt Mitman, Sergio de Ferra and Federica Romei, we investigate heterogeneous transmission of monetary policy among Euro Area countries. We document three empirical regularities across the Euro Area. First, countries with higher inequality have larger service expenditure shares. Second, monetary policy has a weaker output impact in these high-service-share, high-inequality countries. And third, monetary policy induces systematic trade flows between high- and low-service-share countries.

In order to rationalize these findings, we introduce non-homothetic preferences and heterogeneous sectoral income into an otherwise standard New Keynesian model. We show that pro-cyclical income inequality, driven by wealthier households' greater income exposure to services, buffers poorer households' consumption to contractionary shocks and dampens overall policy transmission. These results suggest that accounting for cross-country differences in income and consumption distributions is crucial for understanding common monetary policy.

In the third chapter, titled "Economic Consequences of Large Extraction Declines: Lessons for the Green Transition" and co-authored with Rudolfs Bems, Andrea Pescatori and Martin Stuermer, we study the implications of large declines in resource extraction for fossil fuel exporting economies. Motivated by the imminent green transition, we construct a novel dataset covering the extraction of 13 minerals across 122 countries since 1950, to empirically analyze responses to large, exogenous extraction declines in both domestic and external sectors. We find that such declines lead to persistent reductions in real GDP and the trade balance. Although the real exchange rate depreciates over time, this adjustment is insufficient to offset the deterioration in net exports. Overall, our findings suggest that the green transition may entail substantial economic costs for fossil fuel–exporting countries, particularly for low-income economies with limited capacity to absorb such shocks.

Chapter 1

The Regional Specialization Trade-off

Abstract

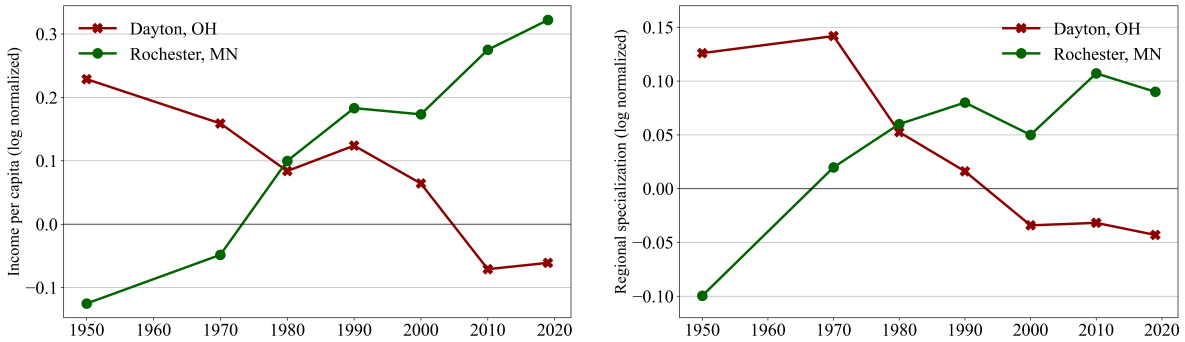
In 1950, highly specialized U.S. regions had higher per capita incomes than those with greater industrial diversity. Since then, however, the more specialized regions have grown persistently slower. I rationalize this novel finding in a dynamic multi-industry model featuring two opposing forces. On the one hand, specialization raises productivity via agglomeration economies. On the other hand, it increases exposure to sectoral shocks. Real factor adjustment costs and financial frictions make reallocation in response to shocks costly and long-lasting. Disciplined by U.S. Census data, the model explains half of the observed relationship between initial specialization and subsequent growth, with financial frictions accounting for more than half of this adverse effect. A constrained-efficient planner allocation reveals that less specialization can raise welfare by reducing a region's exposure to industry-specific downturns.

I am thankful to Federica Romei and Todd Schoellman for their support and guidance. I am also grateful to Andrea Chiavari, Andrea Ferrero, Sergio de Ferra, David Nagy, Fabrizio Perri, and Petr Sedlacek for the helpful comments and discussions.

1.1 Introduction

The economic fortunes of regions can shift dramatically over time. Once-thriving hubs often face long-term decline, while other areas rise to prominence. This paper offers new evidence that a crucial factor in explaining regional variation is the degree of *regional specialization* — the concentration of economic activity in a few specific industries. While economic theory often frames specialization as a benefit of comparative advantage, this paper argues it is a double-edged sword: the industrial structure that makes a region prosperous in the short run can make it vulnerable to steep and lasting declines in the long run.[1]

I begin by documenting three novel facts about U.S. regional growth using U.S. Census microdata. First, commuting zones (henceforth, regions) that were more specialized in 1950 had higher per capita incomes than more diversified regions. Since then, however, these specialized regions have exhibited persistently lower long-run growth. This highlights a *specialization trade-off* where regions benefit from focusing resources on a specific industry today but may be persistently worse off when this industry declines in the future. Second, as regions grow, they become more specialized (and vice versa). Third, the specialization at the region-industry level, however, is highly persistent. Once a region is specialized in one industry, it tends to maintain this specialization in the long run.



(a) Income per capita

(b) Regional Specialization

Figure 1.1: The Specialization Trade-off: Historical Example

Notes: Panel 1.1a shows the relative log per capita income of the Dayton, Ohio and Rochester Minnesota commuting zones relative to the rest of the U.S. since 1950. Panel 1.1b shows the relative regional specialization of the two regions. Specialization is measured as the Gini coefficient on income shares across 3-digit Census industries. Further detail on the measure is provided in section 3.2.

As an example of these facts, consider the divergent economic trajectories of Dayton, Ohio, and Rochester, Minnesota as shown in Figure 1.1. Dayton's economy was once dominated by a single industry - the manufacturing of cash registers, led by the National Cash Register company (NCR). While this specialization initially brought prosperity, the city's fortunes were dependent on this one sector. When technological change and

[1]For a summary of specialization benefits, see e.g. Costinot, Donaldson, et al. (2015).

global competition led to NCR's decline, Dayton's lack of industrial diversity resulted in a prolonged economic descent, leaving a legacy of idle manufacturing plants.

Rochester's development provides a stark contrast. Alongside its mid-century manufacturing base, anchored by an IBM facility, the city fostered a robust healthcare sector centered on the renowned Mayo Clinic. This economic diversity proved critical. When the manufacturing sector declined, the expanding Mayo Clinic provided a new engine for growth. Unlike Dayton, Rochester successfully transitioned its economy, becoming one of the nation's wealthiest mid-size metropolitan areas.

Motivated by these novel empirical facts, this paper examines how regional specialization affects growth and what degree of specialization is optimal. To address these questions, I develop a dynamic multi-sector model that links specialization to growth and formalizes a trade-off between benefits and costs of specialization. In the model, specialization increases productivity and income through industry-specific agglomeration forces. In line with a vast existing literature, these forces capture the *benefits of specialization* (as in e.g., Bartelme et al. (2019)). The model's key contribution is then to endogenize the *costs of specialization* by showing how it can lead to persistently lower long-run growth. These costs arise through two dynamic frictions. First, adjustment costs make reallocating factors after shocks an inherently slow and costly process. Second, a financial friction in the form of an occasionally binding collateral constraint amplifies and propagates adverse shocks, especially for highly specialized regions.[2] Together, these frictions imply that specialization not only increases exposure to industry-specific shocks but also increases the cost of responding to them. For specialized regions this can create "lock-in" effect, where a region's economic trajectory becomes tied to the fate of its dominant industry.

I calibrate the model using U.S. Census data on all 722 commuting zones and use the calibrated model to derive two quantitative results. First, I apply the model to understand U.S. regional growth since 1950. Matching the initial relative income and specialization of all commuting zones and simulating forward using realized industry productivity paths, the model is able to capture 51% of the relationship between relative specialization in 1950 and 1950-2020 income growth across the U.S..

Second, the calibrated model attributes an important role to financial amplification in explaining the persistence of regional specialization and heterogeneity in growth. When a region becomes financially constrained following an adverse sectoral shock, the reallocation of capital across industries is almost halved in the medium term. Intuitively, hitting the collateral constraint limits the amount of financing available for new investment and thereby suppresses the ability to re-specialize the economy. A counterfactual simulation exercise in which a region is not subject to the financial friction shows that

[2]Well-known examples of similar financial amplification include Bianchi and Mendoza (2018) and Kiyotaki and Moore (1997)

more than half of the explained adverse effect of specialization on growth stem from the financial amplification mechanism.

Next, I employ the model to address the second key question of this paper: What is the optimal degree of regional specialization given that future growth across industries is uncertain? To address this, I derive the optimal specialization implemented by a constrained-efficient, regional planner. Two counteracting sources of inefficiencies motivate the planner to deviate from the decentralized, competitive equilibrium allocation. First, the agglomeration externality incentivizes the planner to increase the specialization of a region. This agglomeration effect captures within-industry knowledge and productivity spillovers that increase with industry size and are not internalized by private agents.

The second inefficiency stems from the risk of specialization captured by the occasionally binding collateral constraint. The constraint takes physical capital valued at its market price as collateral. While private agents take this price as given when allocating capital across industries, the planner internalizes the effect of her decisions on the current and future price of capital - the pecuniary externality. I extend the results from standard financial amplification models (Bianchi and Mendoza, 2018; Kiyotaki and Moore, 1997) to a multi-industry setting to show that the responsiveness of the capital price to an industry-specific productivity shock increases with the degree of regional specialization. The risk of becoming financially constrained therefore incentivizes the planner to diversify a region.[3] Intuitively, the planner understands that by being more diversified across industries, an industry-specific shock moves the price of capital by less. As the price of capital determines the tightness of the collateral constraint, a more diversified economy is thereby less likely to become constrained and to enter a vicious downturn.

With these two opposing forces, the optimal degree of specialization becomes a quantitative question. I return to the calibrated model to estimate the socially optimal specialization and measure the welfare difference between the decentralized competitive equilibrium and the constrained-efficient planner equilibrium. The analysis shows that for some regions, the planner increases specialization, as the agglomeration benefits outweigh the risk of industry-specific shocks. Conversely, for others (e.g. Dayton, Ohio), the planner chooses to diversify when the costs of shock exposure are the dominant concern. The model reveals that the industrial composition implemented by a regional planner in 1950 would raise welfare by 1.2 to 2.2 percent depending on the commuting zone. Finally, I show that the planner can implement the optimal specialization using a combination of state-contingent industry taxes and subsidies and a state-contingent tax on debt.[4]

[3] In addition, the planner saves more relative to the private agent as is standard in financial amplification models Bianchi (2011).

[4] The planner will also choose a lower level of bond holdings as is standard in the financial amplification literature.

Related literature. I contribute to three strands of literature. The first one is the growth literature that links regional growth to structural change (see, e.g. Autor and Dorn (2013), Barro and Sala-i-Martin (1992), Francesco Caselli and Coleman II (2001), Eckert and Peters (2022), and Ganong and Shoag (2017)), trade liberalization (Caliendo et al. (2018)), misallocation (Fajgelbaum et al., 2019), employment (Bilal, 2023), innovation (Desmet and Rossi-Hansberg, 2014) and start-up location (Walsh, 2023). Building on this, my primary contribution is to identify and formalize a specialization trade-off in U.S. regional growth. This trade-off demonstrates that the nature of regional economic structure is another important mediating factor of regional convergence.

The notion of specialization trade-offs can be traced back to seminal work by Glaeser et al. (1992), Jacobs (1969), and Jacobs (1985). The most recent, closely related paper is by Heblich et al. (2025). The authors document a specialization trade-off for cities in 19th-century Great Britain and rationalize it using an exogenous, negative externality of regional specialization on growth. My model provides a different mechanism that relies on adjustment costs and financial frictions that generate this effect endogenously.

Second, this paper contributes to the international trade literature by re-evaluating the long-run consequences of specialization (see, e.g. Do and Levchenko (2007) and Stiglitz and Newbery (1984)). A large amount of trade literature portrays specialization according to comparative advantage as a source of welfare gains (Costinot and Donaldson, 2012; Costinot, Donaldson, et al., 2015; Moroney and Walker, 1966). While some studies acknowledge adjustment costs in the short run (Francesco Caselli, Koren, et al., 2020), the long-run benefits of specialization are rarely questioned. I contribute to this field by showing that while specialization can raise short-term income through agglomeration economies (Bartelme et al., 2019; Moretti, 2010), it can also create a lock-in effect that is detrimental to long-run growth, offering a theoretical and empirical counterpoint.

Finally, to capture the risk of the trade-off, I introduce a financial amplification channel, building on the literature on local financial frictions and growth. This work typically examines how shocks are amplified and can lead to financial crises and sudden stops (Bernanke et al., 1999; Gertler and Karadi, 2011; Kiyotaki and Moore, 1997; Mendoza, 2010). My model extends this research by showing how a region's degree of specialization can interact with financial conditions. Specifically, I extend Bianchi and Mendoza (2018) to a multi-industry setting showing how specialized regions may attract concentrated investment in the short run but become more exposed to financial downturns in response to industry-specific shocks in the long run.

The rest of the paper is structured as follows. Section 3.2 provides the empirical evidence. Section 1.3 introduces the model and section 1.4 the quantitative analysis. The constrained-efficient quantitative results are discussed in section 1.5. Finally, section 3.9 concludes.

1.2 Empirical Analysis

I document three novel facts about U.S. regional growth since 1950. Together, these facts show that the degree of regional specialization plays a significant role in determining current income and long-run income growth. Throughout the paper, I will use commuting zones (CZ) as the primary geographic unit which I also refer to interchangeably as a region.

Data. I assemble data from three main sources. First, I collect nationally representative sample of individual-level data from the decennial U.S. Census. The survey is run every ten years between 1950 and 2010. After 2010, I supplement the Census dataset with the American Community Survey (ACS), and the Consumer Population Survey (CPS) together with its Annual Social and Economic Supplement (ASEC). Both cover information on a range of economic, employment, demographic topics including income measures and information on the region of residence. I assign individuals to one of 722 mutually exclusive and exhaustive local labor markets that correspond to commuting zones defined by Dorn (2009). Commuting zones are the most precise way of defining local labor markets. In the analysis that follows, the definition of a local labor market implies that agents cannot move across markets free of costs.

Second, I use aggregate data at the county-industry level for employment, population and GDP across all industries since 1950 from the County Business Patterns (CBP) as well as data on the firm population from the Business Dynamics Statistics (BDS) since 1978. Since the classification of industries changed from SIC to NAICS in 1998, I use the crosswalk from Eckert, Fort, et al. (2020) to match industries over time.

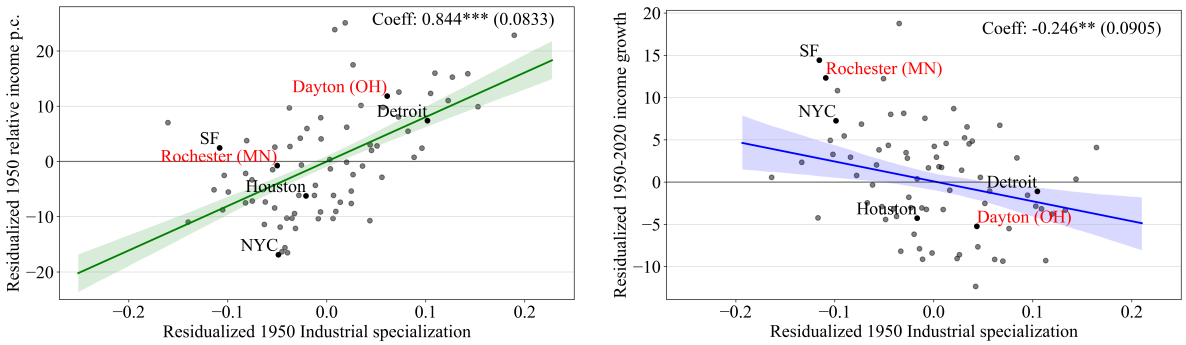
Finally, data on housing and land unit types and values are obtained from the decennial county-level sample of the U.S. Housing Census. The Housing Census runs since 1970 which restricts part of my analysis to the last 50 years.

Measuring specialization. Specialization refers to the concentration of economic activity across industries. Following Imbs and Wacziarg (2003), I measure specialization as the Gini coefficient on economic activity across industries. A Gini coefficient of 1 implies maximum specialization with all value (or employment) generated within one industry. A Gini coefficient of 0 implies full diversification. The baseline measure of industry shares will be pre-tax wage and salary income generated by the working age population (age 25-60) at the 3-digit CPS industry level defined in 1990.[5] I adjust top-coded income variables in line with Heathcote, Perri, Violante, and L. Zhang (2023) and Moffitt and S. Zhang (2018) and run the replacement of top-coded values on state-level rather than U.S.-level income distribution as explained in further detail in Appendix 1.7. I use the U.S. Census 3-digit industry classification which implies 946 industries

[5] Appendix 1.7 shows that the results are robust to choosing value added or employment shares as variables of other indexes (e.g. HHI or max share) as measures for specialization.

at the most granular level. For part of the analysis, I aggregate income shares up to 9 aggregate industries in line with NAICS classification codes. A detailed list of industry details can be found in Appendix 1.7.

Fact 1: *Highly specialized regions are richer in the short-run but have lower long-run growth.*



(a) 1950 Income level

(b) 1950-2020 Income growth

Figure 1.2: The Specialization Trade-off after Controls

Notes: Figure 1.2 shows the residualized linear fit from regression (1.5) for 1950 per capita income level (1.2a) and 1950-2020 per capita income growth (1.2b). Specialization is measured as the Gini coefficient of income shares across 3-digit 1990 U.S. Census industries. 95% Confidence intervals are shaded.

Figure 1.2 presents the specialization trade-off: Highly specialized regions are richer in the short-run but experience lower growth in the long-run. It shows a binned scatterplot of commuting zones and the residualized linear fit between current income (left) and long-run growth (right) and initial 1950 industrial specialization after controls. I estimate this effect across U.S. CZs with the following regression:

$$Y_c = \alpha + \beta \cdot Gini_{c,1950} + \delta \cdot \hat{g}_c + \gamma' \cdot Z_{c,1950} + \epsilon_c \tag{1.1}$$

where Y_c is the dependent variable for CZ c and $Gini_{c,1950}$ is the measure of industrial specialization by 3-digit industry. I augment the regression with two sets of controls. The first accounts for structural change, an important confounding factor. Since 1950, agriculture and manufacturing industries have generally declined, while services industries have expanded. If regions with greater specialization in 1950 were also more concentrated in industries most affected by this structural transformation, then the coefficient β would partly reflect the effects of structural change rather than purely the impact of regional specialization. To isolate the influence of industrial specialization from that of structural change, I follow Borusyak et al. (2025) in using a shift-share control. Specifically, I compute the exposure to structural change by including the predicted growth across industries based on a region's initial share in these industries as follows:

$$\hat{g}_c = \sum_{i=1}^{I} s_{i,c} \cdot g_i \tag{1.2}$$

where $s_{i,c}$ is the 1950 income share and g_i is the 1950-2020 aggregate income growth of industry i. Thus $\hat{g}_c$ captures the predicted growth of region c given its initial exposure to aggregate changes across industries. The coefficient β can then be interpreted as

the effect of specialization on relative income after controlling for exposure to aggregate changes across industries and production structure over time.

The second set of controls $Z_{c,1950}$ includes the 1950 per capita income, population size, share of high-skilled workers, old-age dependency ratio, share of female workers in logs as well as a dummy for whether the CZ is located within the Rust Belt. All of these controls are introduced to capture well-known confounding forces that drive growth between 1950-2020 and may directly relate to the industrial composition of a region.[6]

The regression results indicate that a 1 percentage point higher Gini coefficient of specialization in 1950 is associated with approximately a 0.246 percentage point lower relative income growth over the long run. The full regression specification, including all controls, is reported in Appendix 1.7.

Robustness and Extensions. I run a range of robustness and extension exercises on the regression defined above (see Appendix 1.7). First, I show that higher specialization in 1950 is also associated with a lower income level in 2020. Hence, the slower growth linked to greater initial specialization translates into persistent differences in income levels over time. As illustrated in Figure 1.1a by comparing the evolution of Dayton, Ohio and Rochester, Minnesota, specialization thus helps explain the reversal of economic fortunes across U.S. regions.

Second, it is important to note that the time horizon matters for the relationship between specialization and growth. In the baseline, I choose to define long-run growth over the longest horizon possible given the data to show that specialization matters has a long-lasting impact. In addition, I also run regression (1.5) over shorter time horizons. The results suggest that a negative growth impact of specialization becomes evident over a 20-year horizon already.

Third, I decompose the effect of specialization by industry and tradability. I find that virtually all of the relevance of specialization for contemporary and future relative income of labor markets is linked to specialization within tradable industries. Intuitively, whether a region is disproportionately concentrated within non-tradable industries (e.g. a city has relatively many hairdressers) matters little for relative income and long-run growth. Instead, if all of a region's tradable output is generated in a single industry (e.g. cash registers as in Dayton, Ohio) appears to significantly reduce the likelihood of sustained long-run growth. Moreover, this negative relationship also holds within broad industry groups: regions specialized within manufacturing, agriculture, wholesale, or retail tend to exhibit higher initial income levels but slower long-run growth.

Fact 2: *As regions grow, they become more specialized.*

The second fact captures the dynamics of regional specialization over time. Intuitively, as U.S. regions grow and move up in the income distribution, they become more specialized. And vice versa, regions that once were rich and specialized become more diversified as they decline.

In order to understand the dynamics of regional specialization I proceed in two steps. First, I summarize the evolution of regional specialization across the U.S. over time. Table 1.1 shows the average regional specialization as well as its dispersion in the U.S. since 1950 in terms of both income and employment shares. Since 1950, the U.S. has experienced a rise in regional specialization by 15%. At the same time, the dispersion

[6] A further discussion of these controls can be found in Appendix 1.7.

of specialization has declined by around 50%. Together, these statistics imply that the U.S. is becoming more concentrated both on aggregate and at the regional level.

Year	Income Shares			Employment Shares		
	Mean (1)	CV (2)	p90/p10 (3)	Mean (4)	CV (5)	p90/p10 (6)
1950	0.46	0.162	1.55	0.45	0.152	1.50
1970	0.48	0.123	1.35	0.47	0.116	1.35
1990	0.47	0.089	1.23	0.46	0.082	1.26
2010	0.53	0.089	1.29	0.50	0.073	1.20
2020	0.53	0.089	1.24	0.51	0.068	1.20

Table 1.1: U.S. Regional industrial specialization over time

Notes: Table 1.1 summarizes measures of U.S. regional specialization for Commuting Zones over time. Specialization is measured by the Gini coefficient, calculated on 3-digit industry shares of income and employment. Columns 1 and 4 report the mean specialization across zones. Columns 2 and 5 show the coefficient of variation (CV), and columns 3 and 6 show the 90/10 percentile ratio as measures of dispersion.

Next, I examine how the relative specialization of a region moves relative to the aggregate over time. I log-normalize per capita incomes and Gini coefficients in each decade. The key challenge in characterizing the dynamics of the relationship lies in its potential nonlinearity. Specifically, relative income and specialization may not more monotonically over time. In order to deal with potential nonlinearity, I follow Imbs and Wacziarg (2003) in documenting the relationship between regional sectoral concentration and the relative level of per capita income using a non-parametric locally weighted scatterplot estimation. The idea of this method is to impose as little structure on the functional form of the relationship as possible. The point estimates are computed from fitted values within weighted intervals around the independent variable from the following regression

$$y_i = \alpha(x_i) + \beta(x_i)x_i + \epsilon_i \tag{1.3}$$

where i corresponds to a single observation (Commuting zones - year) and x_i, y_i is the log-normalized income and the Gini coefficient, respectively.[7] The point estimates are then given by

$$(\hat{\alpha}(x_i), \hat{\beta}(x_i)) = \arg\min_{\alpha,\beta} \sum_j w_j(x_i)(y_i - (\alpha + \beta x_j))^2 \tag{1.4}$$

with weights $w_j(x_i)$. [8] The resulting curve of fitted values can be interpreted as the typical degree of industrial specialization for a commuting zone along the income dimension.[9]

Figure 1.3 shows the estimated relationship between relative per capita income and specialization since 1950: The degree of relative specialization increases with relative income along an S-shaped curve. As a region moves within the distribution of per capita incomes in the U.S. it changes its degree of relative specialization. Towards the tail ends of either distribution the relationship flattens indicating that the marginal effect of specialization on income decreases in the extremes. It is important to note, that there is no causality to this statement so it could equally be made in reverse.

[7] For a more detailed explanation see Imbs and Wacziarg (2003).

[8] The estimation uses tricube weights. Note that rectangular or Gaussian weights give similar results.

[9] I also run the estimation for each decade since 1950 with similar estimates across time and as a standard OLS as shown in Appendix 1.7.

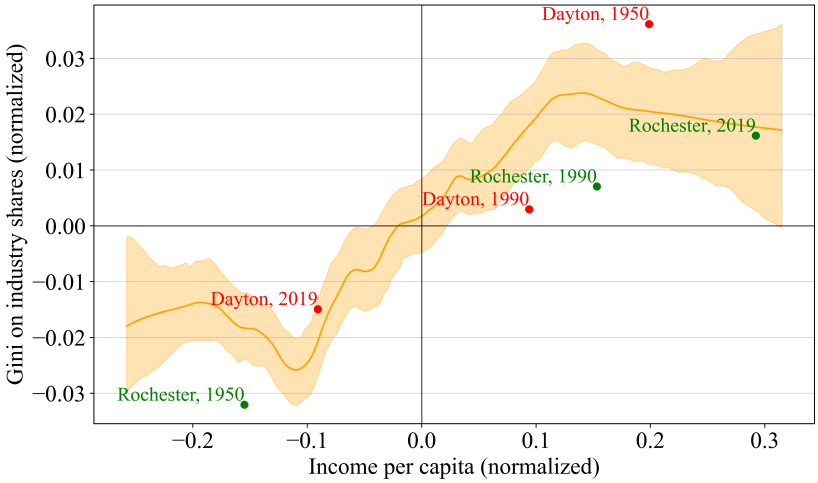


Figure 1.3: Non-parametric estimate: Specialization vs. per capita income

Notes: Figure 1.3 plots fitted values of normalized relative specialization against relative income. The line is estimated using the non-parametric regression in equation (1.3), with the 95% bootstrapped confidence interval shown as a shaded area. The observations for Dayton, Ohio, and Rochester, Minnesota, are highlighted.

Finally, the relationship of specialization and income among U.S. regions appears to differ from the relationship estimated at the cross-country level by Imbs and Wacziarg (2003). The authors find a significant U-shaped relationship where countries first become more diversified and then more specialized as they grow. In their analysis, both the poorest and wealthiest countries are the most specialized while mid-income nations are most diversified.

Fact 3: *Specialization at the region-industry level is highly persistent.*

This fact documents that industries in their relative importance do not reallocate across regions over time. Intuitively, a region that was the biggest producer in a single industry in 1950 (e.g. cash registers as in Dayton, Ohio) is likely to still be the biggest producer in that industry today, irrespective of how that industry developed on the aggregate.

In order to measure the persistence of a region's relative specialization within a certain industry, I define a region's revealed comparative advantage (RCA):

$$RCA_{i,c,t} = \frac{Y_{i,c,t}}{\sum_{i \in I} Y_{i,c,t}} \Big/ \frac{Y_{i,US,t}}{\sum_{i \in I} Y_{i,US,t}} \tag{1.5}$$

where $Y_{i,c,t}$ is the income generated in industry i in CZ c. In line with trade literature, the RCA measures how much more prevalent an industry is in a region relative to the U.S. as a whole. An $RCA > 1$ implies relative specialization. I follow Morris-Levenson (2022) and rank regions based on their RCA in an industry and use this ranking in the following regression:

$$\text{logRankRCA}_{i,c,t} = \alpha + \beta_h \cdot \text{logRankRCA}_{i,c,t-h} + \delta_{c,t} + \gamma_{i,t} + \epsilon_{i,c,t}$$

where β_h is the rank-rank elasticity of the revealed comparative advantage over horizon h. I further control for CZ-year and industry-year effects to account for any structural differences across industries and regions over time. Figure 1.4 shows the persistence (β) of industry rankings over time. The persistence of specialization at the region-industry

level is highest among tradable industries: a 1% increase in the ranking of specializations in industry i in 1950 implies a 0.57% increase in the ranking today. Over 70 years this is a sizable persistence of industries.[10] More detail to persistence across different industries and regions is given in Appendix 1.7.

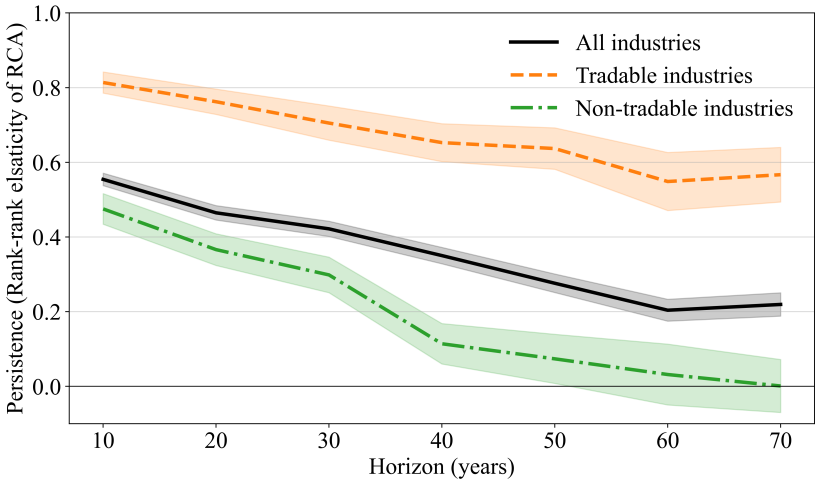


Figure 1.4: Persistence at the region-industry level

Notes: Figure 1.4 shows the rank-rank elasticity of regional revealed comparative advantage over a 70-year horizon, estimated from the regression in equation (1.5). The shaded area represents the 90% confidence interval. The figure also presents separate persistence estimates for tradable and non-tradable industries, as defined in Appendix 1.7.

Combined, Facts 2 and 3 suggest that once a region becomes highly specialized in a single industry, this specialization tends to persist over the long run, tying the region's economic trajectory to the performance of that dominant industry. While this dynamic can be advantageous for regions that happen to specialize in a high-growth sector, the historical evidence suggests that such concentration eventually exposes them to economic decline. This trade-off motivates one of the paper's central questions: What is the optimal degree of regional specialization given that future growth across industries is uncertain? In order to establish the link between specialization and growth theoretically, and to answer this question I propose a model of regional specialization in the following section.

1.3 A Model of Regional Specialization

In this section, I develop a dynamic multi-sector model of specialization and growth. In the model, specialization has two counteracting effects on growth that formalize a trade-off. On the one hand, it increases exposure to industry-specific shocks, making them more costly and long-lasting.

The goal of the calibrated model is then threefold. First, I formalize the interaction between regional specialization, income and growth. Second, I estimate the strength of the proposed mechanisms in generating the specialization trade-off. Finally, I derive the optimal, constrained-efficient specialization a regional planner implements. In this section I present the model, and in the following two sections I discuss the quantitative implications.

[10] Note that Morris-Levenson (2022) find similar estimates over even longer horizons.

1.3.1 Environment

The economy is defined as a small open economy modeled in discrete time, indexed by t. It consists of a production side with multiple industries and a household side with a continuum of individuals.

Production and technology. Production occurs in two stages. A set of intermediate goods is produced by distinct industries, which are then aggregated to create a single final good. Each intermediate good $y_{i,t}$ is produced using capital $k_{i,t}$ in the following production function

$$y_{i,t} = z_{i,t} f(k_{i,t}) \tag{1.6}$$

where $z_{i,t}$ is industry-specific productivity and $f(k)$ is a twice-differentiable, concave production function.

Industries are heterogeneous along three dimensions: productivity, capital adjustment costs and their susceptibility to agglomeration externalities. Industry-specific productivity $z_{i,t}$ is the key driver of dynamics and is assumed to follow a trend-stationary process. This captures both temporary shocks and long-run growth differences across industries. The process is defined as:

$$\bar{z}_{i,t} = \tilde{z}_i + g_{i,t} t + u_{i,t} \tag{1.7}$$

with log of productivity $\bar{z}_{i,t}$ defined as the sum of a linear trend $g_{i,t}$ and a stochastic shock component. The stochastic component follows a standard mean-zero AR(1) process: $u_{i,t} = \rho u_{i,t-1} + \epsilon_{i,t}$ where $\epsilon_{i,t} \sim \mathcal{N}(0, \sigma^2_{\epsilon_i})$.

Capital accumulation in each industry is subject to idiosyncratic convex adjustment costs $\Phi_i(k_{i,t}, k_{i,t+1})$. The adjustment costs introduce persistence in the allocation of factors and represent the costs of tearing down existing machinery and plants or building new infrastructure necessary for new machinery and plants. The difference in costs across industries captures the idea that some capital may be easier to re-use for other industries. Intuitively, heavy manufacturing machinery may be more difficult to demolish than laptops used in the financial sector.

Finally, each industry is subject to an industry-specific agglomeration force:

$$z_{i,t} = e^{\bar{z}_{i,t}} k_{i,t}^{\xi_i} \tag{1.8}$$

where agglomeration parameter $\xi_i \geq 0$ captures productivity spillovers that are not internalized by private agents. In line with Bartelme et al. (2019), these externalities can arise, for example, from knowledge sharing between suppliers, a pooled market for specialized labor, or the common usage of industry-specific infrastructure. They imply that an industry's productivity rises with its size, creating gains from specialization in specific industries.

Individuals. There is a continuum of identical, infinitely-lived individuals of measure unity with preferences given by:

$$\mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t u(c_t). \tag{1.9}$$

Individuals in this economy operate production directly by allocating capital across all industries.[11] They buy and sell capital from each other at the capital market price q_t. In order to invest in new capital, pay the convex adjustment costs and smooth

[11] As shown by Bianchi and Mendoza (2018), the competitive equilibrium is the same if the optimization problems of households and firms are separated (assuming a frictionless equity market).

consumption, individuals can further borrow by selling a one-period risk-free bond b_t with a world-determined gross real interest rate R_t which is taken as given.

The individual's budget constraint is:

$$c_t + \frac{b_{t+1}}{R_t} + q_t \sum_{i=1}^{I} k_{i,t+1} = \sum_{i=1}^{I} \left[z_{i,t} f(k_{i,t}) + q_t k_{i,t} - \Phi_i(k_{i,t}, k_{i,t+1}) \right] + b_t. \tag{1.10}$$

The right-hand side corresponds to the agent's income generated from production in each industry i, selling the beginning-of-period capital stock net of the adjustment costs at price q_t, and beginning-of-period bond holdings. They use this income to consume the final good c_t, buy new capital at price q_t and allocate it across industries, and save in the one-period bond. The total amount of capital is in fixed unit supply.

The total amount of debt individuals can borrow to invest in capital is limited by a collateral constraint.[12] Debt cannot exceed a fraction θ_t of the market value of beginning-of-period asset holdings (i.e. θ_t imposes a ceiling on the leverage ratio). The market value of assets is a combination of the value of capital priced at q_t and the total amount of capital allocated across all industries i. The collateral constraint is given by:

$$-\frac{b_{t+1}}{R_t} \leq \theta_t q_t \sum_{i=1}^{I} k_{i,t} \tag{1.11}$$

The collateral constraint introduces the key financial amplification mechanism capturing the risk of specialization. It extends the results from standard incomplete financial market models (Bianchi and Mendoza, 2018; Kiyotaki and Moore, 1997) to a multi-industry setting. The motivation for this modeling choice is threefold. First, empirical evidence indicates that approximately 70 percent of U.S. commercial and industrial loans are secured by collateral, primarily in the form of fixed capital and real estate (Gan, 2007). Chaney et al. (2012) further show that regional fluctuations in real estate prices significantly influence corporate investment through the collateral channel, explaining much of the cross-regional variation in investment.[13] Second, regions that experience sharp declines in local asset values—such as housing prices or the value of production facilities—face tighter collateral constraints. This tightening reduces borrowing capacity and amplifies the sensitivity of regional consumption to income fluctuations (Caplin et al., 1993; Lustig and Van Nieuwerburgh, 2010). Notable examples include the decline of manufacturing in the U.S. Rust Belt, the collapse of the coal industry in Germany's Ruhrgebiet, and the contraction of cash register manufacturing in the Dayton metropolitan area. Finally, Peek and Rosengren (2000) demonstrate that the collateral channel also operates on the supply side of credit. Banks holding depreciated real estate assets reduce their lending activity, thereby constraining their clients' investment.

The agent maximizes (1.9) subject to (1.10) and (1.11) taking prices as given. This maximization problem yields the following optimality conditions for each date $t = 0, ..., \infty$:

$$u'(c_t)(q_t + \phi^2_{i,t}) = \beta \, \mathbb{E}_t \left[u'(c_{t+1})(q_{t+1} + z_{i,t+1} f'(k_{i,t+1}) - \phi^1_{i,t+1}) + \theta_{t+1} q_{t+1} \eta_{t+1} \right] \tag{1.12}$$

$$u'(c_t) = \beta R_t \, \mathbb{E}_t \left[u'(c_{t+1}) \right] + \eta_t \tag{1.13}$$

[12] In line with Bianchi and Mendoza (2018) this constraint can be derived as an implication of incentive-compatibility constraints on borrowers if limited enforcement prevent lenders from collecting more than a fraction of θ_t of the asset value owned by a defaulting debtor.

[13] In addition, Benmelech and Bergman (2009) find that the redeployability of collateral across industries affects credit spreads: industries with more general (less specific) collateral enjoy higher credit ratings and lower loan-to-value ratios.

where $\eta \geq 0$ denotes the Lagrange multiplier on the collateral constraint, and ϕ^s denotes the derivative of adjustment costs with respect to the s argument. The optimality conditions reveal the two key choices individuals face in this economy: A *portfolio allocation* and a *consumption vs. savings* decision.

Condition (1.12) presents the *portfolio allocation* choice. For each industry i, individuals allocate capital such that the marginal costs of an additional unit equals the marginal benefit. The marginal costs consist of the price of capital today and the marginal adjustment costs weighted by today's marginal utility of consumption. The marginal benefits combine the expected price at which capital can be sold next period net of adjustment costs and the expected marginal product. In addition, each unit of capital adds to the collateral value and loosens next period's collateral constraint. Note that there is only one price of capital q_t clearing the market. Intuitively, in order to sell capital on the market agents first have to turn it into marketable capital by paying the adjustment costs, and then sell it on the market at the single capital price.

The bond Euler equation (1.13) shows the *consumption vs. savings* decision. An unconstrained individual ($\eta_t = 0$) can perfectly smooth consumption across two periods by borrowing or saving in the one-period bond given the real interest rate R_t. When the collateral constraint binds, $\eta_t > 0$, the condition implies that the marginal benefit of borrowing exceeds the expected marginal cost by an amount equal to the shadow price of relaxing the borrowing constraint (i.e. the agent faces an effective real interest rate higher than R_t).

The existence of the occasionally binding collateral constraint introduces a wedge in the standard Euler equations. In the bond Euler equation a binding constraint today impedes consumption smoothing. In the capital Euler equation, the expectation of a binding borrowing constraint next period raises the marginal benefit from capital.

1.3.2 Unregulated Decentralized Competitive Equilibrium

I define and solve for the decentralized equilibrium (DE) in recursive form. I separate individual bond holdings b from the economy's aggregate bond position B on which prices depend. The state variables for the agent's problem are the individual states $\{b, \mathcal{K} = \{k_i\}_{i=\{1,...,I\}}\}$, and the aggregate states $\{B, \mathcal{Z} = \{z_i\}_{i=\{1,...,I\}}\}$. Aggregate capital is not a state variable because it is in fixed supply. In addition, in order to form expectations of future prices, private agents need a "perceived" law of motion $B' = \Gamma(B, \mathcal{K}, \mathcal{Z})$ governing the evolution of the economy's bond position, and a conjectured asset pricing function $q(B, \mathcal{K}, \mathcal{Z})$.

The individual's recursive optimization problem is:

$$V(B, b, \mathcal{K}, \mathcal{Z}) = \max_{c, b', k_i'} \; u(c) + \beta \mathbb{E} V(B', b', \mathcal{K}', \mathcal{Z}') \tag{1.14}$$

$$\text{s.t.} \quad c + \frac{b'}{R} + q(B, \mathcal{K}, \mathcal{Z}) \sum_i^I k_i' = \sum_i^I [z_i f(k_i) + q(B, \mathcal{K}, \mathcal{Z}) k_i - \Phi_i(k_i, k_i')] + b$$

$$-\frac{b'}{R} \leq \theta q(B, \mathcal{K}, \mathcal{Z}) \sum_i k_i$$

$$B' = \Gamma(B, \mathcal{K}, \mathcal{Z})$$

The solution to this problem is characterized by the decision rules $\hat{b}(B, b, \mathcal{K}, \mathcal{Z})$,$\hat{k_i}(B, b, \mathcal{K}, \mathcal{Z})$, and $\hat{c}(B, b, \mathcal{K}, \mathcal{Z})$. The decision rule for bond holdings induces an "actual" law of motion

for aggregate bonds, which is given by $\hat{b}(B, B, 1, \mathcal{Z})$, and the recursive form of (1.16) induces an asset pricing function $\hat{q}(B, 1, \mathcal{Z})$.

Definition 1 (Recursive Competitive Equilibrium)**.** *A recursive competitive equilibrium is defined by an asset pricing functions $q(B, \mathcal{K}, \mathcal{Z})$, a perceived law of motion for aggregate bond holdings $\Gamma(B, \mathcal{K}, \mathcal{Z})$ and the decision rules $\hat{b}(B, b, \mathcal{K}, \mathcal{Z})$, $\hat{k}_i(B, b, \mathcal{K}, \mathcal{Z})$ $\forall\, i$, and $\hat{c}(B, b, \mathcal{K}, \mathcal{Z})$ with associated value function $V(B, b, \mathcal{K}, \mathcal{Z})$ such that:*

1. *The functions $\left\{\hat{b}(B, b, \mathcal{K}, \mathcal{Z}), \hat{k}_i(B, b, \mathcal{K}, \mathcal{Z}), \hat{c}(B, b, \mathcal{K}, \mathcal{Z})\right\}$ and $V(B, b, \mathcal{K}, \mathcal{Z})$ solve the agent's maximization problem, taking as given $q(\cdot)$ and $\Gamma(\cdot)$ as given.*
2. *The market for capital clears: $\sum_i \hat{k}_i(B, b, \mathcal{K}, \mathcal{Z}) = 1$*
3. *The resource constraint holds: $\hat{c}(B, b, \mathcal{K}, \mathcal{Z}) + \frac{\hat{b}(B,b,\mathcal{K},\mathcal{Z})}{R} = \sum_i [z_i f(\hat{k}_i(B, b, \mathcal{K}, \mathcal{Z}))$ $-\Phi_i(\hat{k}_i(B, b, \mathcal{K}, \mathcal{Z}), \hat{k}_i(B', b', \mathcal{K}', \mathcal{Z}'))] + B$*
4. *The perceived and actual laws of motion are consistent: $\Gamma(B, \mathcal{K}, \mathcal{Z}) = \hat{b}(B, B, 1, \mathcal{Z})$ and $\hat{q}(B, \mathcal{K}, \mathcal{Z}) = q(B, \mathcal{K}, \mathcal{Z})$.*

1.3.3 The Role of Specialization

Specialization in this economy has two key roles: determining the agglomeration strength, and determining the exposure to and amplification of industry-specific shocks. The impact of specialization on agglomeration is straightforward and in line with existing literature (Bartelme et al., 2019). Given a a set of agglomeration parameters across industries ξ_i, a more specialized economy exhibits higher productivity and higher income per capita at every point in time.

To capture the exposure to and amplification of sectoral shocks, two frictions generate the costs of specialization. First, standard real factor adjustment costs make reallocation in response to shocks inherently costly and long-lasting. Second, the financial friction in form of the occasionally binding collateral constraint amplifies and propagates adverse shocks, especially for highly specialized regions.

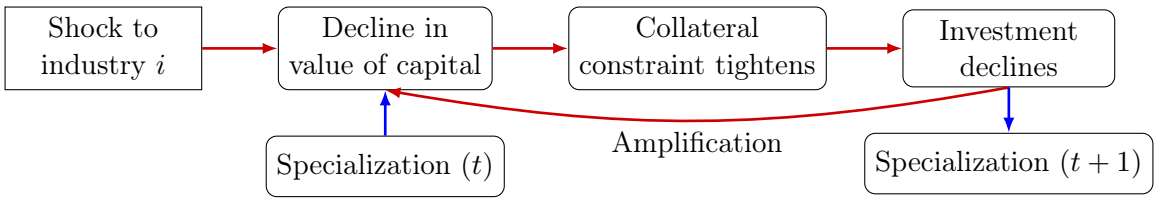


Figure 1.5: Illustrative interaction between specialization and the financial friction

The interaction of specialization with the price of capital and the collateral constraint is the key novel feature of this model. Figure 1.5 illustrates this interaction. In the standard, single-industry financial amplification model (Kiyotaki and Moore, 1997) an adverse productivity shocks reduces the value of productive capital. This tightens the collateral constraint and reduce investment and consumption. The result is an amplification mechanism whereby the lower investment further reduces the value of capital which in turn further tightens the constraint. In a multi-industry setting, specialization interacts with this mechanism in two ways. First, the beginning-of-period specialization affects the sensitivity of the price of capital to industry-specific shocks. Second, becoming financially constrained affects the ability to invest and adjust specialization. The financial friction can thus result in a lock-in effect such that highly specialized regions are more likely to become financially constrained and less able to adjust their capital

allocation in the future.

Specialization and the value of capital. The link between the current capital allocation and the value of capital can be illustrated by extending standard conditions for asset pricing. I proceed in two steps. First, derive the aggregate, weighted dividend as follows:

$$\bar{d}_{t+1} = \sum_{i}^{I} s_{i,t}(z_{i,t+1}f'(k_{i,t+1}) - \phi^1_{i,t+1}) \tag{1.15}$$

where the capital share in industry is $s_{i,t} = \frac{k_{i,t}}{\sum_i k_{i,t}}$. Equation (1.15) shows that the aggregate dividend is the weighted average marginal product of capital net of adjustment costs across all industries. Specialization in this economy can be defined as the Gini coefficient on industry shares $\omega_t \equiv Gini(s_{i,t})$.

Next, re-writing the definition of asset returns in aggregate terms $R^{\text{agg}}_{t+1} = \frac{q_{t+1}+\bar{d}_{t+1}}{q_t+\bar{\phi}_t}$, I express the pricing condition (1.12) as the expected present value of dividends discounted with R^{agg}_{t+1}:

$$q_t = \mathbb{E}_t \sum_{j=0}^{\infty} \mathbb{E}_t \left[\prod_{s=0}^{j} (R^{\text{agg}}_{t+1+s})^{-1} \right] \bar{d}_{t+j+1} - \bar{\phi}^2_t \tag{1.16}$$

where $\bar{\phi}_t = \sum_i s_{i,t}\phi_{i,t}$ are today's weighted aggregate adjustment costs.[14] Equation (1.16) shows that the price of capital today depends on the current and expected allocation of capital that determine the discount rate R^{agg}, the industry-specific productivity shocks that determine the flow of future dividends $\bar{d}_{t+j+1}$, and the size of adjustment costs ϕ_i.

The allocation of capital determines the co-movement of the price of capital with industry-specific productivity shocks. Greater specialization increases the exposure of the value of capital to industry-specific shocks. This interaction can be highlighted considering three illustrative cases:

1. *Standard financial amplification ($I = 1$, No adjustment costs)*: With only one industry and no adjustment costs, the model collapsed to the standard incomplete markets model (Kiyotaki and Moore, 1997; Mendoza, 2010). In this case, the price of capital is determined only by the dividends of the single industry $\bar{d}_t = d_{i,t} = z_{i,t}f'(k_{i,t})$.

2. *Diversification with many industries ($I > 1$, No adjustment costs)*: With many industries and no adjustment costs, the co-movement of the asset price and idiosyncratic productivity shocks tends to zero, and the price of capital to unity: $\lim_{I\to\infty} Cov\,(q, z_i) = 0$. All idiosyncratic risk is diversified away as capital can freely be moved across industries. Intuitively, as soon as an industry is hit by an adverse shock, all capital will be reallocated to another industry and the value of capital is unrelated to the dividend in that particular industry.

3. *Specialization matters ($I > 1$, Positive adjustment costs)*: With many industries and positive adjustment costs, the current capital allocation matters for the co-movement of the capital price and industry-specific shocks: $Cov\,(q, z_i) = F(s_i)$. In this case, the price of capital is determined by the current and expected allocation

[14] Likewise, the pricing condition can also be written as $q_t = \mathbb{E}\sum_{j=0}^{\infty} \beta^j \frac{u'(c_{t+j})}{u'(c_t)} \bar{d}_{t+j+1} - \mathbb{E}\sum_{j=0}^{\infty} \beta^j \frac{u'(c_{t+j})}{u'(c_t)} \bar{\phi}^2_t$.

of capital across industries. In this economy, an industry-specific shock becomes an aggregate shock since capital cannot be freely moved.

The third case illustrates the relevant key relationship in this paper: The more diversified an economy, the less exposed it is to industry-specific productivity shocks and the less its price of capital co-moves with the productivity of a single industry. The above mechanism is at the core of the model's pecuniary externality that will be further discussed in the planner's section.

The collateral constraint and the ability to invest. Now consider the second part of the feedback loop: the interaction between the collateral constraint and next period's capital allocation. To study the relationship, I re-write the bond Euler equation as:

$$R_t \beta \mathbb{E}_t \Big[\frac{u'(c_{t+1})}{u'(c_t)} \Big] = R_t \mathbb{E}_t[\mathcal{M}_{t,t+1}] = (1 - \eta_t) \tag{1.17}$$

where $\mathcal{M}_{t,t+1}$ denotes the stochastic discount factor (SDF). Condition (1.17) shows how the collateral constraint affects subjective discounting of individuals. When the constraint is not binding, the SDF equals one and intertemporal substitution is based one the world interest rate R_t. The more the constraint binds, the more η exceeds zero, the greater the subjective discounting. Intuitively, the more constrained an agent is today, the more she wants to move resources to today and the less she cares about the future.

Plugging the above into the capital Euler equation (1.12) shows the impact of the collateral constraint on the portfolio allocation:

$$q_t = \beta \, \mathbb{E}_t \Big[\frac{(1 - \eta_t)}{R_t} (q_{t+1} + z_{i,t+1} f'(k_{i,t+1}) - \phi^1_{i,t+1}) + \frac{\theta q_{t+1} \eta_{t+1}}{u'(c_t)} \Big] + \tau_{i,t} - \phi^2_{i,t} \tag{1.18}$$

where $\tau_{i,t} = \beta Cov(\mathcal{M}_{t,t+1}, (q_{t+1} + z_{i,t+1} f'(k_{i,t+1}) - \phi^1_{i,t+1}))$. Condition (1.18) shows that becoming collateral constrained today acts as a wedge $(1-\eta_t)$ in the portfolio allocation. Individuals discount expected marginal benefits of capital allocation more the more they are constrained today. As a result, a constrained agent will invest less into reallocating capital and is stuck with her beginning-of-period industrial specialization. The covariance term $\tau_{i,t}$ captures the additional adjustment due to risk: assets whose payoffs move with marginal utility (i.e., high in bad states) command higher prices, while risky assets are discounted more heavily.[15]

In combination, the two parts of the interaction between specialization and the financial friction imply a novel feedback loop. First, highly specialized economies are more likely to become financially constrained in case of adverse productivity shocks Second, financially constrained economies are less able to change their initial specialization. As a result, highly specialized regions may become locked into their dominant industries and decline alongside them. This effect is particularly severe when shocks are persistent, since financial constraints hinder the capacity to re-specialize into new, more viable industries.

1.3.4 Constrained-efficient Planner Equilibrium

In this section, I consider the optimal policy of a constrained-efficient social planner (SP). The concept of constrained efficiency describes the best feasible allocation taking

[15] When sectoral shocks are largely idiosyncratic, the covariance term $\tau_{i,t}$ is small, so the constraint wedge $(1-\eta_t)$ remains the dominant channel through which collateral scarcity affects portfolio allocation.

market incompleteness as given. The planner in this scenario cannot undo the market incompleteness and erase the financial constraint but maximizes aggregate welfare given this constraint by accounting for externalities and instructing private agents to deviate from their private optimality conditions.[16]

The SP chooses allocations while internalizing the two key externalities in the economy: the agglomeration externality and the pecuniary externality stemming from the collateral constraint. The SP is subject to the same collateral constraint as the private agents (i.e. he respects that the capital price remains market-determined). In addition to the private agents, however, he understands how the capital price entering the constraint is determined by taking the capital Euler conditions as additional implementability constraints. Let $\left\{\hat{b}(B,b,\mathcal{K},\mathcal{Z}),\hat{k}_i(B,b,\mathcal{K},\mathcal{Z}),\hat{c}(B,b,\mathcal{K},\mathcal{Z})\right\}$ be the policy functions that return the values of the corresponding variables in the competitive equilibrium. Taking these functions as given, the SP's optimization problem can be represented in recursive form as follows:

$$
\begin{aligned}
V^{SP}(b,\mathcal{K},\mathcal{Z}) &= \max_{c,b',k_i',q} \; u(c)+\beta\mathbb{E}V^{SP}(b',\mathcal{K}',\mathcal{Z}') \qquad (1.19)\\
\text{s.t.} \quad c+\frac{b'}{R} &= \sum_{i=1}^{I}\left[z_i f(k_i')-\Phi_i(k_i,k_i')\right]+b\\
-\frac{b'}{R} &\leq \theta q\\
u'(c)(q+\phi_i^2(k_i,k_i')) &= \beta\,\mathbb{E}_t\Big[u'(\mathcal{C}(b',\mathcal{K}',\mathcal{Z}'))(\mathcal{Q}(b',\mathcal{K}',\mathcal{Z}')+z_i'f'(k_i')\\
&\qquad -\phi_i^1(k_i',\mathcal{K}_i(b',\mathcal{K}',\mathcal{Z}')))+\theta'\mathcal{Q}(b',\mathcal{K}',\mathcal{Z}')\eta(b',\mathcal{K}',\mathcal{Z}')\Big]\;\forall i\\
\sum_{i=1}^{I}k_i &= \bar{K}\\
z_i &= \bar{z}_i\cdot k_i^{\xi_i}\quad\forall\, i
\end{aligned}
$$

The economy's resource constraint has a multiplier $\lambda \geq 0$, the collateral constraint $\eta^* \geq 0$ which is different from the private η as private and social values from relaxing the collateral constraint may differ. The implementability constraints for each industry i have multipliers $\delta_i \geq 0$, and the market clearing Lagrange multiplier is denoted by Ξ.

Definition 2 (Recursive Constrained-efficient Equilibrium)**.** *The recursive constrained-efficient equilibrium is defined by the policy functions $k_i(b,\mathcal{KZ})$, and $b(b,\mathcal{KZ})$ with associated decision rules $\hat{c}\,(b,\mathcal{K},\mathcal{Z})$, $\eta^*\,(b,\mathcal{K},\mathcal{Z})$, the pricing function $q\,(b,\mathcal{K},\mathcal{Z})$ and value function $V\,(b,\mathcal{K},\mathcal{Z})$, and the associated decision rules $\mathcal{C}(b,\mathcal{K},\mathcal{Z})$, $\eta(b,\mathcal{K},\mathcal{Z})$ and asset prices $\mathcal{Q}(b,\mathcal{K},\mathcal{Z})$ such that the following conditions hold:*

1. *Planner's optimization: $V\,(b,\mathcal{K},\mathcal{Z})$ and the functions $b(b,\mathcal{KZ})$, $k_i(b,\mathcal{KZ})$, $\hat{c}\,(b,\mathcal{K},\mathcal{Z})$, $\eta^*\,(b,\mathcal{K},\mathcal{Z})$ and $q\,(b,\mathcal{K},\mathcal{Z})$ solve the Bellman equation defined in Problem (1.19) given $\mathcal{C}(b,\mathcal{K},\mathcal{Z})$, $\eta(b,\mathcal{K},\mathcal{Z})$ and $\mathcal{Q}(b,\mathcal{K},\mathcal{Z})$.*

Efficiency and optimal policy. Efficiency in this economy can be characterized by the difference between the planner and private agents' solutions. This difference can be summarized by two key components. On the one hand, the existence of the agglomeration externality which private agents fail to internalize. On the other hand, the effect of current capital allocation (i.e. portfolio allocation) and bond holdings (i.e. consumption

[16] For further discussion of constrained efficiency see, e.g. Korinek (2011).

vs saving) on the likelihood of becoming financially constrained via the price of capital and the existence of the collateral constraint. I refer to the latter as the pecuniary externality.

I characterize these differences by defining the following planner optimality conditions:

$$c_t :: \lambda_t = u'(c_t) - \frac{u''(c_t)}{u'(c_t)}\theta_t \eta_t^* q_t - u''(c_t) \sum_{i=1}^{I} \delta_{i,t} \phi_{i,t}^2 \tag{1.20}$$

$$q_t :: \sum_{i=1}^{I} \delta_{i,t} = \frac{\theta_t \eta_t^*}{u'(c_t)} \tag{1.21}$$

$$k_{i,t+1} :: \lambda_t \phi_{i,t}^2 = \beta \mathbb{E}_t \Big[\lambda_{t+1} (\bar{z}_{i,t+1} (\xi_i + \alpha) k_{i,t+1}^{\xi_i + \alpha - 1} - \phi_{i,t+1}^1) + \sum_{i=1}^{I} \delta_{i,t} \Omega_{i,t+1}^K \Big] + \Xi \tag{1.22}$$

$$b_{t+1} :: \lambda_t = \beta R_t \mathbb{E}_t \Big[\lambda_{t+1} + \sum_{i=1}^{I} \delta_{i,t} \Omega_{i,t+1}^B \Big] + \eta_t^* \tag{1.23}$$

where Ω^B and Ω^K collect all terms with derivatives that capture the effects of the SP's choice of b' and k_i' via all implementability constraints. They are further discussed in Appendix 1.8. The socially optimal policy can be defined in terms of the social *value of wealth*, *portfolio allocation* and *consumption vs savings* choice.

Condition (1.20) presents the social *value of wealth*. It extends the private value, which takes the standard form $\lambda_t = u'(c_t)$, by two terms. The second term on the right-hand side corresponds to the amount by which an additional unit of consumption reduces marginal utility and relaxes the collateral constraint by raising the asset price. This term is standard in incomplete financial market models (Bianchi and Mendoza, 2018). Note, that the term disappears when the collateral constraint is currently non-binding, and $\eta_t^* = 0$. The third term on the right-hand side reflects the amount by which an additional unit of consumption reduces marginal utility and relaxes each industry's implementability constraint at the adjustment costs ϕ_i. The relaxation of each implementability constraint depends on how strongly each constraint binds which is summarized by the shadow price δ_i. Condition (1.21) shows that the social benefit from relaxing the implementability constraints is positive if and only if the collateral constraint binds. Conversely, it the constraint is non-binding for the social planner, $\sum_{i=1}^{I} \delta_{i,t} = 0$, there is not additional social benefit from raising the price of capital through an extra unit of c_t.

Condition (1.22) presents the socially optimal *portfolio allocation*. The planner allocated capital to each industry such that the social marginal cost equal the expected social marginal benefit in each industry. It deviates from the private agents' allocation in three ways. First, the planner internalizes industry-specific agglomeration economies. Consequently, the social marginal product of capital in industry i exceeds the private marginal product by the term ξ_i. Second, the planner internalizes the pecuniary externality: the effect of capital allocation on the equilibrium price of capital. This is captured by the term $\sum_{i=1}^{I} \delta_{i,t} \Omega_{i,t+1}^K$. Intuitively, the planner understands that shifting capital between industries alters its price, which in turn affects the tightness of future borrowing constraints. The shadow value of these constraints is captured by the Lagrangian multipliers embedded in $\Omega_{i,t+1}^K$. This mechanism introduces a portfolio allocation channel for financial amplification. As a result, the economy's degree of specialization (not just its aggregate bond holdings or marginal utility of consumption) becomes a determinant of its vulnerability to financial shocks. Finally, the planner takes into account the market

clearing condition via the Lagrangian multiplier Ξ.

Condition (1.23) governs the social planner's optimal consumption-savings decision. Unlike private agents, the planner internalizes the pecuniary externality that aggregate debt imposes on the price of capital. This effect, captured by the term $\sum_{i=1}^{I} \delta_{i,t}\Omega^B_{i,t+1}$, reflects how an additional unit of debt today tightens future borrowing constraints. Intuitively, higher debt reduces the economy's flexibility to adjust its capital structure and moves it closer to the collateral limit, thereby increasing the likelihood of a binding constraint. This mechanism extends the one identified in Bianchi (2011) by incorporating the additional channel of capital allocation. Furthermore, the planner's valuation of the collateral constraint, represented by the Lagrangian multiplier η^*, differs from the private multiplier because the social value of relaxing the constraint is not fully internalized by individual agents.

I now show that the planner's equilibrium can be decentralized with a state-contingent tax on debt τ^B_t and a state-contingent industry-specific tax or subsidy on marginal product of capital τ^K_t. I refer to the latter as industrial policy. The price of bonds becomes $\frac{1}{R(1+\tau^B_t)}$ in the budget constraint of the private agent in the regulated competitive equilibrium. The marginal return on capital becomes $(1+\tau^K_{i,t})z_{i,t}k^{\alpha+\xi_i-1}_{i,t}$. The tax revenue is rebated using a lump-sum transfer T_t. The agent's Euler conditions for bonds and capital become:

$$u'(c_t)(q_t+\phi^2_{i,t}) = \beta\,\mathbb{E}_t\left[u'(c_{t+1})(q_{t+1}+(1+\tau^K_{i,t+1})z_{i,t+1}f'(k_{i,t+1})-\phi^1_{i,t+1}) \right.$$
$$\left. +\theta_{t+1}q_{t+1}\eta_{t+1}\right] \tag{1.24}$$

$$u'(c_t) = \beta R_t(1+\tau^B_t)\,\mathbb{E}_t[u'(c_{t+1})] + \eta_t. \tag{1.25}$$

Proposition 1 (Decentralization with taxes and subsidies)**.** *The constrained-efficient equilibrium can be decentralized with a state-contingent tax on debt, and an industry-specific tax (subsidy) with tax revenue rebated as a lump-sum transfer and the tax rates set to satisfy:*

$$1+\tau^B_t = \frac{1}{\mathbb{E}_t[u'(t+1)]}\Big[\sum_i \delta_{i,t}\Omega^B_{i,t+1} - \frac{u''(t+1)}{u'(t+1)}\theta\eta_{t+1}q_{t+1} - u''(t+1)\sum_i \delta_{i,t+1}\phi^2_{i,t+1}\Big]$$
$$+\frac{1}{\beta R_t\mathbb{E}_t[u'(t+1)]}\mathbb{E}_t\Big[\frac{u''(t)}{u'(t)}\theta\eta_t q_t + u''(t)\sum_i \delta_{i,t}\phi^2_{i,t}\Big] \tag{1.26}$$

$$1+\tau^K_{i,t} = \frac{1}{\mathbb{E}_t[\alpha z_{i,t+1}k^{\alpha+\xi_i-1}_{i,t+1}]}\Big[\Big(u'(t+1) - \frac{u''(t+1)}{u'(t+1)}\theta\eta_{t+1}q_{t+1} - u''(t+1)\sum_i \delta_{i,t+1}\phi_{i,t+1}\Big)$$
$$((\alpha+\xi_i)z_{i,t+1}k^{\alpha+\xi_i-1}_{i,t+1} - \phi^1_{i,t+1}) + \sum_i \delta_{i,t}\Omega^K_{i,t+1}$$
$$- u'(t)(q_{t+1}-\phi^1_{i,t}) - \theta\eta_{t+1}q_{t+1}\Big]$$
$$+\frac{1}{\beta\mathbb{E}_t[\alpha z_{i,t+1}k^{\alpha+\xi-1}_{i,t+1}]}\Big(u'(t)q_t + \Xi + \phi^2_{i,t}u''(t)\sum_i \delta_{i,t}\phi^2_{i,t}\Big) \tag{1.27}$$

where the arguments of the functions have been shorthanded as dates to keep the expression simple. The proof can be found in Appendix 1.8.

There are three sources of inefficiencies the planner aims at alleviating with the optimal tax schedules. On the one hand, the planner addresses the pecuniary externality stemming from debt. Private agents fail to internalize the effect of their aggregate bond

holdings on the future price of capital, q_{t+1}, and thus on the tightness of future collateral constraints. The state-contingent tax on debt, τ_t^B, corrects this over-borrowing externality. By comparing the agent's Euler equation for bonds (1.25) with the planner's condition (1.23), it becomes evident that τ_t^B is set to precisely offset the wedge between the private and social valuation of debt, forcing agents to internalize the social costs captured by the $\sum_{i=1}^{I} \delta_{i,t}\Omega_{i,t+1}^B$ term.

On the other hand, two distinct inefficiencies plague the allocation of capital. First, a technological externality arises as agents do not account for industry-specific agglomeration economies, ξ_i. Second, another pecuniary externality exists because agents overlook how their portfolio choices today affect future asset prices through the $\sum_{i=1}^{I} \delta_{i,t}\Omega_{i,t+1}^K$ term. The industrial policy, $\tau_{i,t}^K$, is an instrument that corrects both distortions simultaneously. Note, that this instrument is state- and industry-specific. Hence, the optimal policy will generally involve subsidies for some industries and taxes for others. It adjusts the private marginal return on capital to reflect its true social marginal product, which incorporates both the agglomeration benefits and the general equilibrium effects on financial constraints. By appropriately setting this tax schedule, the planner aligns the private portfolio condition (1.24) with the social optimum (1.22). In this way, the combination of these two policy instruments decentralizes the planner's allocation, restoring efficiency by instructing private agents to deviate from their private optimality conditions.

1.4 Quantitative Analysis

In this section, I study the model's implications by conducting simulations for a baseline calibration. I then use the model to perform three exercises, which together quantify the contribution of regional specialization to growth. First, I apply the model to understand U.S. regional growth since 1950 and rationalizing the specialization trade-off documented empirically. Second, I show the impulse responses in capital portfolio allocation and bond holdings to industry-specific productivity shocks to illustrate the role of the financial amplification mechanism. Finally, I characterize how a planner would have changed regional specialization in 1950 and highlight the welfare implications.

1.4.1 Calibration

I calibrate the model using regional U.S. Census and BEA data between 1950 and 2020 at a decadal frequency. Due to data availability I will predominantly focus on industries within manufacturing for the years before 1987 and use less frequent data for more industries only in a descriptive sense. After 1987, I supplement the manufacturing data with data on all industries. For some variables, I will use only aggregate data because of data limitations. The functional forms for preferences and adjustment costs are the following:

$$u(c_t) = \frac{c_t^{(1-\gamma)} - 1}{(1-\gamma)} \quad \gamma \geq 0$$

$$\Phi_i(k_{i,t}, k_{i,t+1}) = \frac{\Phi_i}{2}\left(\frac{k_{i,t+1}}{k_{i,t}} - 1\right)^2 k_{i,t} \quad \Phi_i \geq 0 \ \forall i$$

with coefficient of risk aversion γ and the adjustment cost parameter Φ_i.

The calibration proceeds in two steps. First, a subset of parameter values are set using direct empirical evidence or standard values from the literature. Second, the remaining

parameters are determined by solving the model to jointly target moments from the data.

In the first step, I set the parameters of the productivity processes and the values of $\{\gamma, \alpha, \{\Phi_i, \xi_i\}_{i=\{1,\dots,I\}}, \theta\}$, the parameters of the R process and the industry-specific TFP processes. The relative risk aversion (or inverse inter-temporal elasticity of substitution) γ is an important parameter because it affects the magnitudes of price adjustments as well as the curvature of the stochastic discount factor when the collateral constraint binds. For a given η, an increase in γ will raise stochastic discounting and thereby reduce the portfolio allocations of individuals. In choosing the risk aversion parameter, I aim to accommodate the estimates of two relevant strands of literature. On the one hand, the real business cycle literature uses a risk aversion of around $\gamma = 2$ (Mendoza, 2010). On the other hand, the finance and asset pricing literature estimates risk aversion between 7.5 and 10 (Bansal et al., 2016). In order to accommodate both literature I choose a baseline risk aversion of $\gamma = 5$ and perform robustness exercises for the boundary values. The income share of capital α is set to be 0.3 reflecting the average historical U.S. capital income share in line with vast existing literature.

The values for industry-specific convex capital adjustment costs are taken jointly from Hall (2004) and Groth and Khan (2010). The former paper estimates adjustment costs from capital Euler equations using U.S. firm-level data over the period 1948 through 2001, and the latter provides updated estimates until 2010. Estimates are reported at the 2-digit SIC industry level of manufactured goods. The detailed table of Φ_i is shown in Appendix 1.7. The calibration of this parameter reveals two important aspects. First, industries vary significantly in their capital specificity and adjustment costs (e.g. fabricated metal industries have more than 50% larger adjustment costs than wood products). It is therefore important to account for these differences in the multi-industry setting of the model. Second, more than one third of all industries have adjustment costs that are close to (or indifferent from) zero at the annual level. In these industries real factor adjustment costs can therefore only play a minor role in generating persistence of capital allocation.

The values for industry-specific agglomeration strengths are taken directly from the estimates reported by Bartelme et al. (2019) and are shown in Appendix 1.7. The authors estimate sector-level scale elasticities using predicted demand as instruments based on trade shares, population and sector size at the 2-digit SIC level. Data limitation allows them to base their estimates only on the years since 1995. In this paper, I therefore assume that agglomeration externalities have remained constant over time. I further assume agglomeration economies to be constant across space such that a single industry generates the same agglomeration irrespective of whether it is located in, for example, Dayton, Ohio or Rochester, Minnesota.

The credit regime value θ is set to be consistent with the average U.S. aggregate corporate leverage over time reported by Graham et al. (2015). The authors construct a century-long panel of U.S. firm data showing that the loan-to-value (LTV) ratio of firms has fluctuated relatively little in the $\theta = [0.3, 0.4]$ range. Their estimates coincide with other studies in the financial amplification literature targeting the frequency of crises (see, e.g. Bianchi (2011) and Gertler and Karadi (2011)). In the baseline calibration, I assume this value to be constant over time.

The industry-specific TFP follows independent, trend-stationary AR(1) processes given by equation (1.7). The shocks $\epsilon_{i,t}$ are discretized using Tauchen's quadrature method with 15 realizations for each shock. I use the 2-digit manufacturing industry-specific

TFP measure following Becker et al. (2021) for the 1958 - 2018 period, the longest time series available from official sources. In order to estimate the processes, I follow a two-step procedure. First, I normalize all TFP by the weighted-averaged TFP across all industries. Then, I estimate relative HP-filtered cyclical and trend components for each industry as given in Appendix 1.7.

The gross interest rate follows an independent AR(1) process.[17] In line with standard approach in the international macro literature Bianchi and Mendoza (2018), I calibrate the interest-rate process by measuring the annualized ex-post return on 90-day U.S. T-bills from the official FRED source. This yields $\bar{R} = 1.013$ with persistence $\rho_R = 0.01$ and variance $\sigma_R = 0.0186$.

The second stage of the calibration is to jointly set the values of $\{\{k_{i,1950}\}_{i\in\{1,\dots,I\}}, \beta$. I invert the model given the externally calibrated capital share α, the agglomeration ξ_i and the 1950 industry TFP $z_{i,1950}$ in order to back out the initial capital stocks (i.e. the initial specialization) for every commuting zone. The target for setting the value of β is an estimate of the average net foreign asset position (NFA) as a share of GDP at 1976 - the earliest available official NFA estimate. In line with existing literature, I do not target the time series average because the U.S. NFA-GDP ratio has displayed a marked downward trend since the early 1980s due to the Global Imbalances phenomenon (Bianchi and Mendoza, 2018). The model's decentralized equilibrium yields an unconditional mean of b as a share of GDP that matches the NFA-GDP ratio with $\beta = 0.95$.

Parameter	**Value**	**Source or Target**
Parameters set externally		
Risk Aversion	$\gamma = 5$	Average value in literature
Capital Share	$\alpha = 0.3$	Avg. U.S. capital income share
Adjustment Costs	$\Phi_i \in [0, 3.26]$	Hall (2004); Groth (2010)
Agglomeration	$\xi_i \in [0.1, 0.29]$	Bartelme et al. (2024)
Collateral regime	$\theta = 0.35$	Historical LTV ratio (Graham et al, 2015)
Interest Rate	$\bar{R} = 1.3\%, \rho_R = 0.01$ $\sigma_R = 0.0186$	U.S. 90-day T-Bills
TFP Process	$\rho_i \in [0.71, 0.9]$ $\sigma_i \in [0.013, 0.027]$	Std. and autoc. of U.S. industry TFP
Parameters set internally		
1950 capital stock	$k_{i,1950} \in [0.1, 0.29]$	Matching income shares
Discount Factor	$\beta = 0.95$	Avg. NFA position = -20% of GDP

Table 1.2: Model Calibration

The model is solved using a global, nonlinear solution algorithm taking into account the occasionally binding, stochastic collateral constraint. The DE solution is obtained using a time iteration algorithm. Following Bianchi (2011), in the SP's problem, I use a nested fixed-point algorithm: The inner loop solves for policy functions and the outer one updates future policies given the solution to the Bellman equation. The existence of multiple, independent productivity processes requires an additional state variable

[17] This assumption is in line with the observation that the Basu-Fernald U.S. Solow residual estimates are uncorrelated with the U.S. real interest rate on 90-day Tbills.

capturing the state of relative productivities at every point in time. Appendix 1.8 describes the algorithm in further detail.

1.4.2 The Specialization Trade-off

I first apply the calibrated model to understand U.S. regional growth since 1950. To investigate the mechanisms through which specialization determines income and long-run growth, I conduct a simulation of the universe of U.S. commuting zones, treating each as an individual small open economy. The simulation spans a 70-year period, initialized using the observed 1950 specialization patterns and driven by the realized productivity processes observed across industries. For the purposes of this exercise, I make several simplifying assumptions.

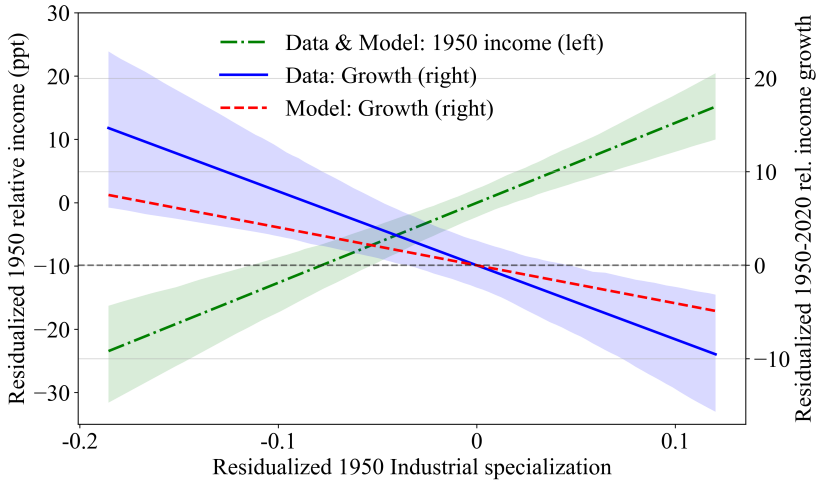


Figure 1.6: The specialization trade-off: Data vs Model

Notes: Figure 1.6 compares the model's simulated effect of specialization on long-run growth (red dotted line) with the observed empirical data. The 70-year simulation is initialized to exactly match the specialization patterns and income levels of U.S. commuting zones in 1950.

First, the industrial landscape is condensed into two sectors: industrial machinery and a composite rest of manufacturing. By focusing on specialization within the manufacturing sector, this approach abstracts from broader structural change issues between manufacturing and other parts of the economy. This is a reasonable simplification, as the main empirical section explicitly controls for structural change using a shift-share control. Second, the rest of manufacturing sector serves as a proxy for a diversified industrial base, representing a safe portfolio of industries. Specialization is hence the concentration of capital across the two sectors. Finally, the simulation is calibrated to directly match the initial specialization shares and resulting income levels for every commuting zone in 1950, ensuring the model's starting conditions match the historical data exactly.

I compare the growth across the economies as a result of the initial specialization and the endogenous re-specialization in response to productivity changes. Figure 1.6 shows that the model can explain roughly half of the long-run growth decline implied by 1950 specialization. The figure further rationalizes the two key aspects of the specialization trade-off. First, greater specialization implies higher short-run income. In the model this is the result of both industry-specific agglomeration and initial capital allocation across industries. The model matches this exactly by default as given industry-specific productivity, agglomeration strength and capital income shares, the initial capital allocations is calibrated to match the observed per capita income differences.

Second, greater initial specialization results in lower long-run growth. This is the result of two mechanisms. On the one hand, the difference in long-run productivity growth across industries in combination of the distribution of initial specialization and the capital adjustment costs creates a mechanical difference in growth rates. On the other hand, the endogenous amplification of the industry-specific shocks through the collateral constraint generates an additional response based on initial industrial specialization.

Figure 1.7 decomposes the effect of specialization on long-run growth by the two frictions. The financial friction generates 56% of the effect of specialization on growth. In absence of the financial frictions, the convex adjustment costs alone capture less than one fourth of the heterogeneity in growth based on specialization. This result highlights the crucial role of financial amplification in explaining long-run growth outcomes, a channel whose quantitative importance this paper underscores. In contrast, the minor effect of adjustment costs aligns with the findings in existing literature (e.g., Hall (2004)) that point towards small empirically estimated adjustment costs across industries.

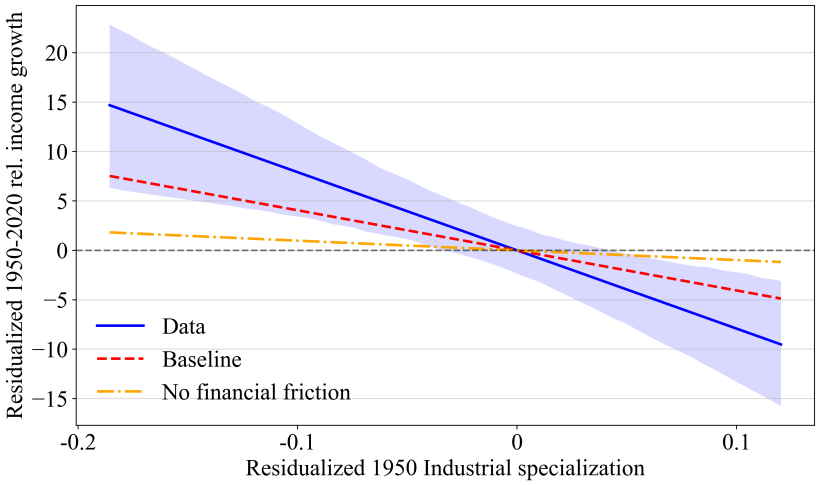


Figure 1.7: The role of financial amplification for growth

Notes: Figure 1.7 decomposes the effect of specialization on 70-year income growth. It compares the relationship observed in the data with two simulations: the baseline model (which includes both financial frictions and adjustment costs) and a counterfactual model without financial frictions.

1.4.3 Specialization Dynamics

In order to analyze the dynamics of regional specialization, I now consider the impulse response to an adverse, one standard deviation transitory shock to a single industry. Figure 1.8 shows the responses of bond holdings, capital allocation to the shocked industry, consumption and the capital price for an economy with and without the collateral constraint beginning from a stationary equilibrium. Two central messages arise from this exercise.

First, the dynamics of the shock are hump-shaped and propagated. In response to the decline of productivity, agents experience a reduction in income. They borrow in order to allocate capital away from the shocked industry and in order to smooth consumption over time. The price of the capital falls as future discounted dividends decline with the marginal product of capital.

Second, comparing the constrained and the unconstrained economy highlights the relevance of the collateral constraint in amplifying and propagating the shock. While the unconstrained economy has unlimited borrowing capacity, the constrained economy faces

a financial amplification that affects both capital allocation and consumption smoothing ability. IN response to the shock, borrowing agents hit the constrained/ This has two immediate effects. First, the standard financial amplifier effect: in order to consume more today, agents start fire-selling capital which reduces the price of capital further and further tightens the borrowing constraint. The amplifier results in the difference between the unconstrained and constrained economies in the bond holdings, consumption and capital price responses. Second, the capital allocation amplifier: constrained agents are less able to invest in capital and pay the re-allocation costs. With greater stochastic discounting they care less about the marginal benefits of re-allocating capital for tomorrow and show a much more sluggish response in capital allocation. The collateral constraint thus implies that agents are stuck with their initial specialization and do invest as much to benefit from capital in more productive industries. This is the key novel driver of persistent industrial specialization.

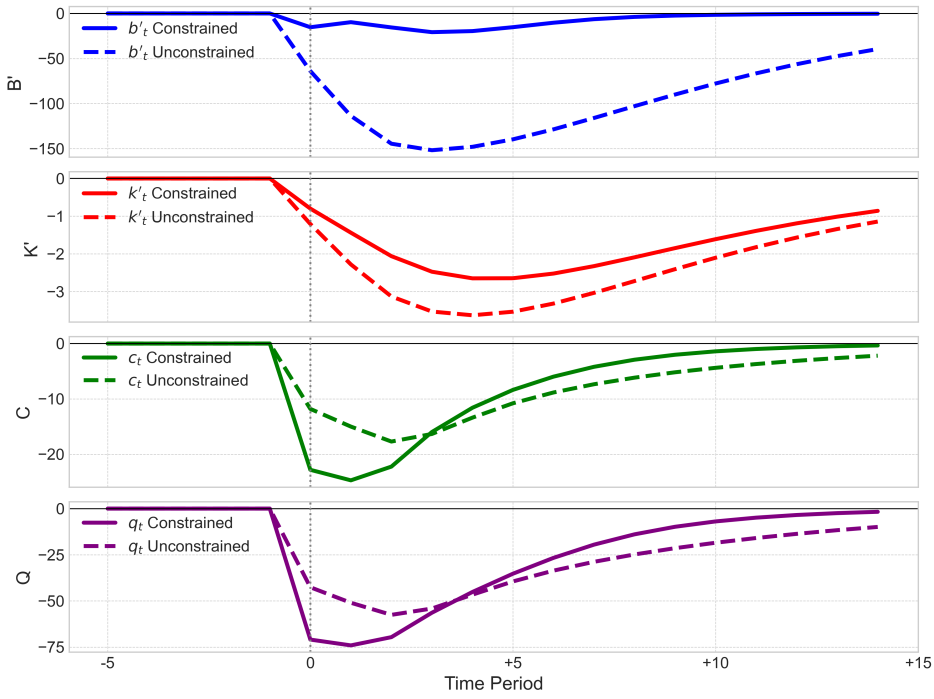
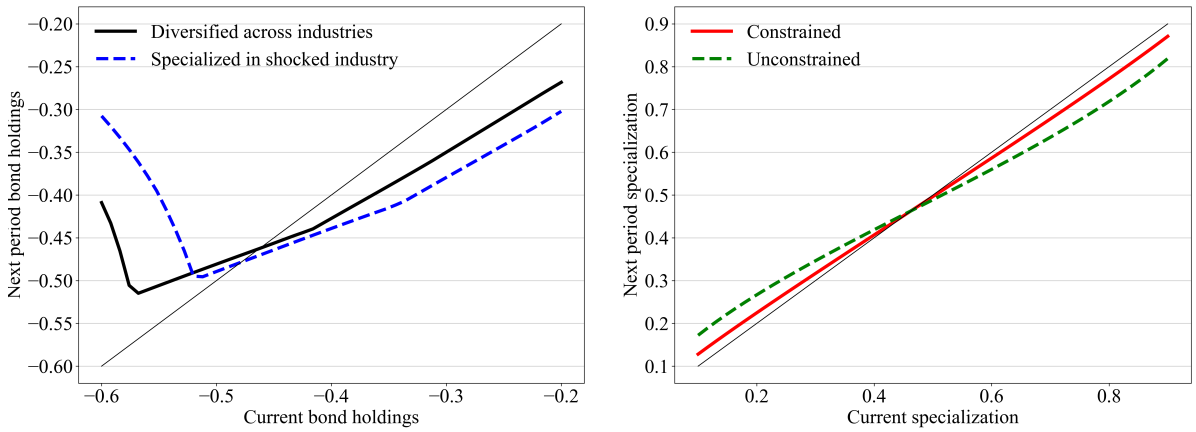


Figure 1.8: Impulse Responses to a transitory industry-specific productivity shock
Notes: Figure 1.8 shows the impulse response functions to a one-standard-deviation adverse productivity shock in a single industry. It compares the response in the baseline model to a counterfactual model without financial frictions.

1.4.4 Capital Allocation, Borrowing Decisions and Amplification

In this section, I solve for the competitive bond and capital accumulation decisions. Figure 1.9a shows the bond policy for a productivity shock to one industry for an economy that is relatively specialized in that industry (e.g. Dayton, Ohio) vis-à-vis an economy that has a diversified portfolio of capital across industries (e.g. Rochester, Minnesota). The graph reveals two key mechanisms. First, without the endogenous borrowing constraint, the policy function for next period's bond holdings would be monotonically increasing in current bond holdings. In line with the standard financial amplification mechanism (Bianchi, 2011), the policy functions are nonmonotonic. The change in the sign of the slope indicates the point at which the credit constraint binds. To the right of this point, the credit constraint is slack and bond decision rules display the usual upwards-sloping shape. To the left of this point, the financial amplifier is at play, where

a fall in the price of capital tightens the constraint forcing a fire-sale of capital which further reduces the price and further tightens the borrowing constraint.



(a) Bond holdings

(b) Capital allocation

Figure 1.9: Policy functions for a negative one-standard deviation industry-specific shock
Notes: Figure 1.9 shows the decision rules for bond holdings and capital allocation following a one-standard-deviation adverse productivity shock to a single industry.

The second mechanism is evident when comparing the two lines and highlights the role of specialization for borrowing decisions and captures the key novel features. The economy specialized in the shocked industry will be more exposed to the shame shock and therefore try to borrow more for any current bond position. Given the stronger fall in the capital price, this economy will hit the borrowing constraint for a higher initial bond position. Intuitively, the more specialized a economy, the more likely it is to become borrowing constraint in response to a shock to that industry. This mechanism extends the standard financial amplification mechanism to a multi-industry setting.

Figure 1.9b shows the capital allocation decision rules: the current share of capital allocated to the industry shocks is given by the horizontal axis, and tomorrow's share of capital in the same industry on the vertical axis. The decision rules are plotted for economies: one that is borrowing constraint and one that is unconstrained. Again, the graph can be interpreted in two steps. First, observe that the slope of the decision rules is smaller than 45 degrees. Hence, in response to an adverse shock to one industry, agents re-allocate capital away from that industry towards the rest of the economy. Second, the ability to re-allocate capital towards other industries is directly affected by the tightness of the borrowing constraint. A constrained economy will be less able to buy new capital and pay the re-allocation costs than an unconstrained economy. This is a combination of the lower liquidity of constraint agents as well as the greater stochastic discounting of any future benefits of capital allocation shown in (1.18).

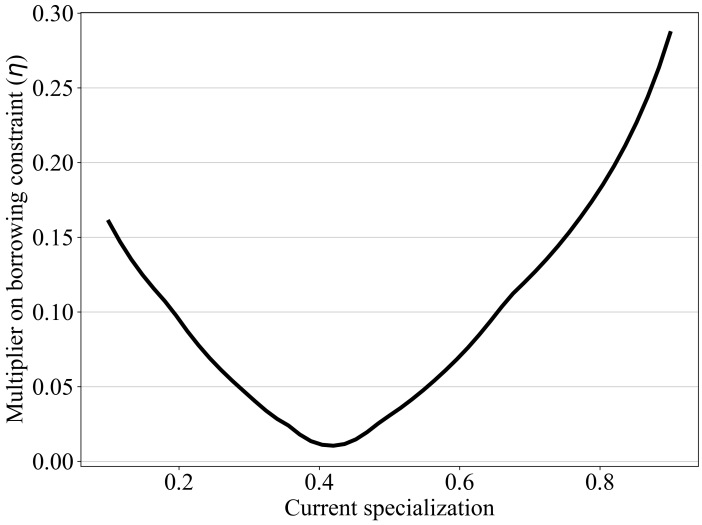


Figure 1.10: Specialization and the tightness of the collateral constraint

Notes: Figure 1.10 Lagrangian multiplier on the collateral constraint over current bond holdings for an adverse one standard deviation productivity shock to a single industry.

The role of current specialization for the likelihood of becoming borrowing constraint is summarized in Figure 1.10. The figure shows the Lagrangian multiplier on the collateral constraint as a function of current capital allocation in the shocked industry. The resulting line depicts a U-shaped curve. The low-point of the curve corresponds to a diversified economy in which capital is evenly allocated across the two industries (i.e. Machinery and the rest of manufacturing). To the right of this point, the tightness of the collateral constrain increases with the degree of specialized in the shocked industry. The fact that the tightness also increases to the left of this points reflects the forward-looking behavior of agents and the capital price. Intuitively, agents that are relatively specialized in the non-shocked industry understand that given their expectation of industry-specific shocks they are over-specialized in the other industry. This understanding is incorporated in the discounting for future expected dividends and leads to a reduction in the capital price today already which tightens the borrowing constraint.

The combination of the bond and capital allocation decision rules capture the novel propagation and financial amplification mechanism of industrial specialization in this model. The feedback loop consists of two parts. On the one hand, greater industrial specialization increases exposure to industry-specific shocks and thereby raises the likelihood of becoming borrowing constraint. On the other hand, borrowing constraint agents are less able to invest into new capital and change their current portfolio of capital across industries. This generates persistence of specialization for highly specialized regions that are hit by an adverse shock to their main industry by suppressing investment. At the same time it generates amplification to negative shocks and introduces an asymmetry in the response to productivity fluctuations.

1.5 Constrained-efficient Allocation and Welfare

In this section, I address the normative question of this paper: What is the optimal degree of regional specialization? I compute the optimal capital allocation and borrowing decision implemented by a constrained-efficient planner that I have in section 1.3. I then proceed in two steps. First, I show how and why the planner equilibrium differs from the competitive one. Then, I present the welfare difference across the two equilibria.

Two sources of inefficiency motivate the planner to deviate from the competitive equilibrium. First, a positive agglomeration externality incentivizes more specialization. The planner internalizes that the social return to concentrating capital exceeds the private return, as individual agents fail to account for how their investments collectively boost industry-level productivity. This motivates the planner to favor a higher degree of industrial concentration than arises in the decentralized equilibrium.

Second, a pecuniary externality, stemming from financial frictions, creates a countervailing incentive for diversification. Individual agents do not internalize how specializing makes the aggregate value of collateral more volatile and exposed to sectoral shocks. The planner recognizes this heightened volatility increases the probability of future borrowing constraints becoming binding across the economy. Consequently, the planner favors diversification as a form of insurance to mitigate this systemic financial risk.

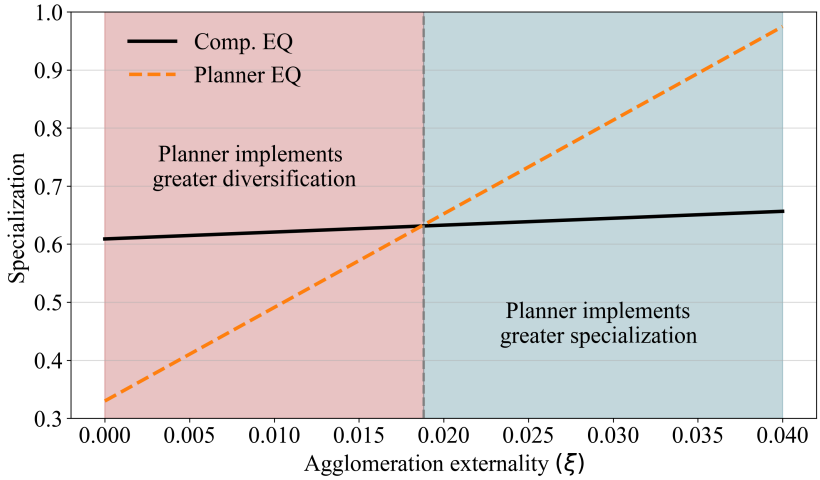


Figure 1.11: Competitive vs constrained-efficient specialization trade-off

Notes: Figure 1.10 compares the specialization in the competitive equilibrium and the constrained-efficient equilibrium for different levels of agglomeration forces in a single industry.

Figure 1.5 summarizes the difference between the socially optimal and private specialization trade-off. I show the regional specialization relative to the strength of agglomeration calibrated using the 1950 level of industry productivities. The graph reveals three key aspects. First, in absence of agglomeration, the competitive equilibrium is over-specialized in too few industries. Private agents fail to internalize the financial amplification to sectoral shocks stemming from the pecuniary externality, and expose the economy too much to becoming borrowing constraint. Instead, the planner will diversify the economy to limit the likelihood of hitting the collateral constraint in response to adverse shocks.

Second, as agglomeration forces strengthen, specialization increases in both equilibria, but more steeply in the planner's allocation. While private agents respond only to the private productivity gains, the planner also internalizes the positive social returns from agglomeration, creating a stronger incentive to concentrate capital. This leads to a unique crossover point where the two opposing inefficiencies—the negative financial externality and the positive agglomeration externality—perfectly offset each other. At this point, the competitive equilibrium is coincidentally socially optimal.

Third, beyond the crossover point, the agglomeration externality dominates the financial risk. The social benefits of concentrating production now outweigh the costs of financial fragility. Consequently, the planner's equilibrium becomes more specialized than the competitive equilibrium. This outcome aligns with the vast literature in economic ge-

ography and trade where specialization is typically found to be welfare-improving (see, e.g. Bartelme et al. (2019)).

Figure 1.12 geographically illustrates the divergence between the constrained-efficient planner's allocation and the competitive equilibrium's specialization for U.S. commuting zones in 1950. The map displays the difference between the degree of specialization a social planner would implement, given the initial conditions of each region, and the level resulting from the decisions of private agents. Commuting zones are categorized into three groups, each representing a tertile of this distribution.

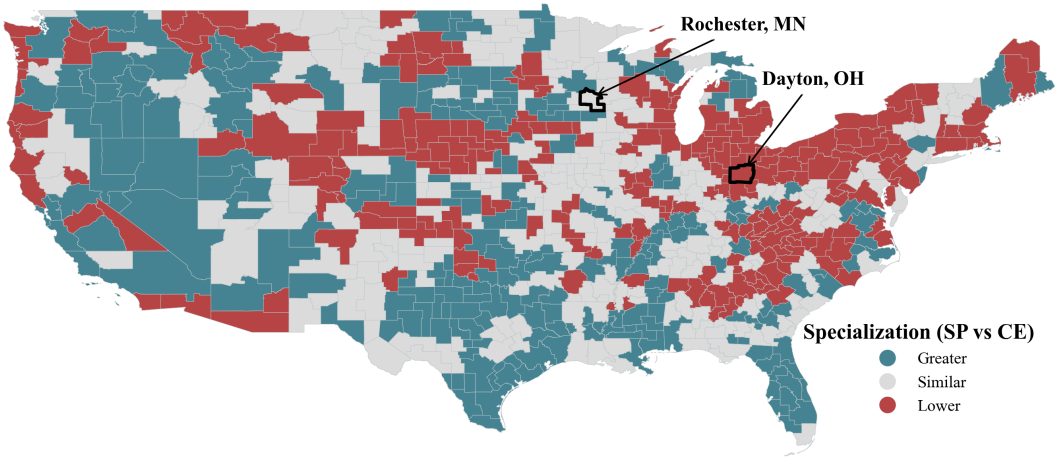


Figure 1.12: Competitive vs constrained-efficient specialization in 1950

Notes: Figure 1.12 compares the socially optimal level of specialization with the competitive equilibrium outcome across U.S. commuting zones in 1950.

The first group (Lower) identifies regions where the planner would enforce greater economic diversification, indicating they were over-specialized in the competitive equilibrium. Geographically, these regions in 1950 were concentrated in the Rust Belt, the lower East, and the West Coast. These cases correspond to the area left of the crossover point in Figure 1.5, where the negative financial externality is the dominant concern. The second group (Similar) includes regions where the planner's allocation is broadly similar to the market outcome. In the third group (Greater), the planner would promote a higher degree of specialization than private agents, suggesting these regions were under-specialized; many of these were located in Florida and Texas. This group corresponds to the area to the right of the crossover point in Figure 1.5, where the positive agglomeration externality outweighs the financial risk.

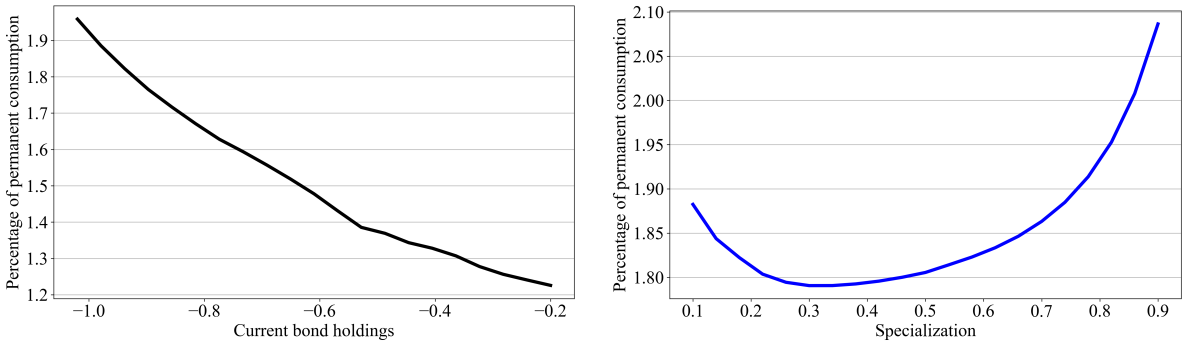
To further illustrate these outcomes, the figure highlights the two specific examples introduced earlier: Dayton, Ohio, and Rochester, Minnesota. From a constrained-efficient perspective, Dayton was over-specialized in 1950, implying an excessive exposure to idiosyncratic sectoral shocks and a heightened risk of financial distress. Conversely, Rochester, Minnesota, appears to have had a degree of specialization closer to the planner's optimum, reflecting a more balanced trade-off between the benefits of concentration and the costs of financial fragility.

Welfare. In line with Bianchi (2011), I compute welfare gains from correct the two opposing externalities as the proportional increase in consumption for all possible future histories in the decentralized equilibrium that would make private agents indifferent between remaining in the decentralized equilibrium and correcting the externalities. Because of the homotheticity of the utility function, the welfare gains Λ at a state

$(b, \sum_i k_i)$ is given by:

$$(1 + \Lambda(b, \sum_i k_i)^{(1-\gamma)}) V^{DE}(B, b, \sum_i k_i) = V^{SP}(b, \sum_i k_i) \tag{1.28}$$

Figure 1.13 quantifies the welfare gains from the planner's optimal policy following a one-standard-deviation adverse productivity shock. The gains are evaluated across the economy's initial state, defined by its asset position and degree of specialization. Panel 1.13a shows that welfare gains are greatest for the most indebted economies. Intuitively, regions with large negative bond holdings are more financially fragile and operate closer to their borrowing limit. The planner's policy provides insurance by encouraging higher savings (i.e., holding more bonds) and reallocating capital to reduce the probability of the borrowing constraint binding after a shock. This intervention is most valuable for economies initially on the brink of financial distress.



(a) Bond holdings

(b) Capital allocation

Figure 1.13: Welfare gains for a negative one-standard deviation industry-specific shock
Notes: Figure 1.13 shows the welfare gain of the constrained-efficient planner allocation over current bond holdings and specialization.

Panel 1.13b reveals a U-shaped relationship between the welfare gains and the initial level of specialization. Gains are substantial in both tails. For highly specialized regions, the planner corrects the competitive equilibrium's failure to account for systemic financial risk. For highly diversified regions, the planner pushes for more concentration to reap the benefits of agglomeration externalities that private agents ignore. The gains are minimized for regions where the initial level of specialization happens to be near the optimal trade-off between agglomeration benefits and financial risk, the point where the two externalities largely offset each other, as illustrated previously in Figure 1.11.

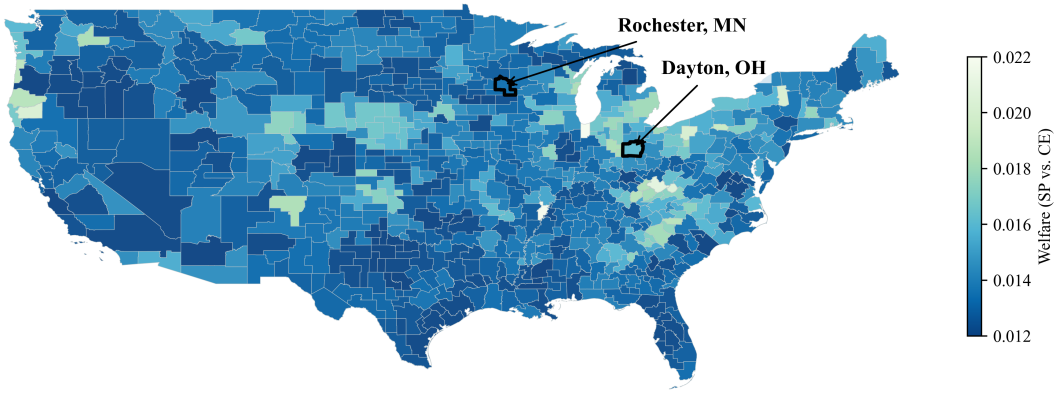


Figure 1.14: Welfare under the constrained-efficient specialization in 1950

Notes: Figure 1.12 shows the welfare gain under the constrained-efficient allocation in 1950 as percentage of permanent consumption.

Finally, Figure 1.14 quantifies the welfare implications of adopting the constrained-efficient allocation. The gains, measured in terms of equivalent permanent consumption, are economically significant, ranging from 1.2 to 2.2 percent depending on regional characteristics. The geographical distribution of these gains is highly uneven and directly reflects the initial misallocation of capital. The most substantial welfare improvements are concentrated in the over-specialized regions of the Rust Belt and the lower East. For these areas, the planner's policy of diversification provides crucial insurance against sectoral shocks, thereby reducing financial fragility and yielding large welfare benefits. Conversely, the gains are more modest in under-specialized regions like Texas and Florida, where the planner's intervention primarily corrects for the agglomeration externality, leading to smaller, though still positive, welfare improvements.

1.6 Conclusion

In this paper, I show that industrial specialization of regions is crucial for understanding regional income and growth differences. For my analysis, I use regional growth differentials across 722 U.S. commuting zones since 1950. I document three novel facts. First, commuting with high industrial specialization in 1950 have higher per capita income initially but experience persistently lower growth in the long-run. This highlights a specialization trade-off where regions benefit from focusing resources on a specific industry today but may be persistently worse off if this industry declines. Second, as regions grow, they become more specialized (and vice versa). Third, the specialization at the region-industry level, however, is highly persistent.

Given this empirical evidence, I develop a dynamic multi-region, multi-sector model to link specialization to growth and formalize the trade-off between benefits and costs of specialization. The model's key innovation is to endogenize the costs of specialization by showing how it can lead to persistently lower long-run growth. The costs arise through adjustment costs and a financial friction. The quantitative model, disciplined by U.S. Census data on commuting zones, reveals that more than half of the adverse effect of specialization on growth can be attributed to the financial friction. A constrained efficient planner allocation that internalizes the benefits of specialization through agglomeration externalities, and the costs of specialization through the financial frictions highlights the potential for industrial policy. By balancing the two opposing force when allocating factors, the planner increases welfare by 1.2 to 2.2 percent depending on the region.

This paper offers the potential vast future research, including cross-regional spillovers of specialization and aggregate growth implications of regional specialization.

1.7 Appendix - Empirical Analysis

Top-coding CPS data

Top-coding is an important issue to address in the CPS for measuring income at the top of the income distribution. I follow Heathcote, Perri, and Violante (2010) in dealing with top-coded observations by assuming the underlying distribution of income is Pareto, and follow as suggestion from David Domeij to extrapolate top-coded observations based on this assumption. Since regional variation is the relevant factor in this paper, I deviate from the authors by defining the distribution at the state-level rather than the country-level. A more detailed explanation of the top-coding process is given by Heathcote, Perri, and Violante (2020).

Industry details

Table 1.3 details the employment shares of 17 major industries for the years 1950, 1990, and 2020. These aggregate sectors are constructed from 865 detailed 3-digit industry classifications sourced from the U.S. Decennial Census data. The table further categorizes each industry as either tradable or nontradable. This distinction is based on a qualitative assessment of whether an industry's output primarily competes within a local market or on a broader national scale. Tradable industries are considered those whose output can be readily sold and consumed far from its point of production; this sector includes industries like Agriculture, Mining, and nearly all forms of Manufacturing. Nontradable industries, conversely, primarily serve local demand with output that is consumed locally; this sector is composed mainly of services, such as Construction, Retail Trade, Education, and Personal Services. This conceptual distinction is useful for contextualizing the patterns of specialization documented in the main text.

	Industry	1950	1990	2020	Tradable
1	Agriculture	20.71	3.61	3.46	Yes
2	Business Services	2.96	4.43	7.61	Yes
3	Communication	0.61	1.52	1.36	No
4	Construction	8.75	9.98	11.91	No
5	Durable	13.53	15.88	10.77	Yes
6	Entertainment	0.66	1.06	1.28	No
7	Finance	2.20	4.47	4.79	No
8	Mining	3.99	1.90	1.82	Yes
9	Nondurable	9.48	8.64	5.77	Yes
10	Personal Services	2.37	1.39	1.60	No
11	Routine Prof. Serv.	4.39	11.26	13.19	No
12	Non-routine Prof. Serv.	0.37	2.02	3.33	Yes
13	Public	4.67	7.96	6.98	No
14	Retail	11.84	11.15	13.26	No
15	Transportation	8.09	6.61	6.91	Yes
16	Utilities	1.80	2.53	2.34	No
17	Wholesale	3.59	5.60	3.63	Yes

Table 1.3: Employment shares

Fact 1: Robustness and Extensions

In this section, I provide a range of robustness and extension exercises on Fact 1 as described in section 3.2.

Regression Table: Equation (1.5)

Table 1.4 shows the regression results from equation (1.5). The control variables are the following. Specialization is given by the Gini coefficient on income shares by 3-digit CPS industry. I decompose specialization further into tradable and non-tradable industries. ll other variables are included at their 1950 level. The exposure to structural change is captured by the shift-share instrument $\hat{g}$. High-skill share is the share of employed workers with at least one year of higher education. The old-age dependency ratio is the share of workers younger than 25 over workers older than 25.

The results presented in Table 1.4 provide a detailed empirical foundation for Fact 1. Columns (1), (3) and (5) directly test the specialization trade-off highlighted in the main text. The coefficient on the overall specialization is negative for long-run growth (-0.246), while being positive for initial 1950 income per capita (0.844). On average, more specialized commuting zones in 1950 were initially richer but experienced slower growth over the subsequent 70 years. Moreover, this relationship between 1950 specialization and economic outcomes remains even when considering 2020 income levels. Regions that were more specialized in 1950—particularly in tradable industries—tend to have lower income per capita in 2020, confirming the long-run persistence of the specialization effect.

Columns (2) and (4) extend this finding by decomposing specialization into its tradable and nontradable components as defined in Table 1.5. This decomposition reveals that the trade-off is driven almost exclusively by specialization in tradable industries. A higher concentration in tradable sectors is associated with slower long-run growth (-0.151) and higher initial income (0.110). In contrast, the coefficients for specialization in nontradable industries are insignificantly different from zero, suggesting that concentration in local services and similar sectors does not exhibit the same trade-off. Intuitively, whether a region has disproportionately many hairdressers cannot predict a regions expected growth and relative income. If, however, all tradable output is generated in a single tradable sector (e.g. cash registers as in Dayton, Ohio) appears to be a relevant predictor for future growth.

Across all four specifications, the control variables are largely significant and informative. The highly significant negative coefficient on the initial log income p.c. in the growth regressions provides strong evidence of conditional convergence, indicating that poorer regions tended to grow faster. Human capital, measured by the high-skill labor share, is a powerful predictor of both higher initial income and faster subsequent growth. Similarly, the 1950 population is positively correlated with both outcomes, consistent with the presence of agglomeration economies. The shift-share instrument for structural change, $\hat{g}$, exhibits a negative coefficient across all models, suggesting that regions with an industrial mix that was predicted to grow based on national trends tended to have lower initial incomes and experience slower growth, once other local characteristics are controlled for. Overall, the models demonstrate substantial explanatory power, with an adjusted R-squared of approximately 0.54 for the full growth models.

	1950-2020 Growth		1950 Income p.c.		2020 Income p.c.	
	(1)	(2)	(3)	(4)	(5)	(6)
Specialization	-0.246**		0.844***		-0.229*	
	(0.0905)		(0.084)		(0.101)	
Tradable		-0.151***		0.110**		-0.00412
		(0.0388)		(0.0423)		(0.0436)
Non-tradable		0.632*		-0.43		-0.122
		(0.294)		(0.356)		(0.357)
1950 measures:						
$\hat{g}$	-0.180***	-0.133*	-0.187**	-0.319***	0.00101	0.0379
	(0.0471)	(0.0521)	(0.058)	(0.0631)	(0.0535)	(0.0674)
log income p.c.	-0.868***	-0.885***			0.168**	0.136***
	(0.0339)	(0.0323)			(0.0439)	(0.0401)
High-skill labor share	1.435***	1.618***	6.042***	5.692***	0.869	1.199*
	(0.41)	(0.418)	(0.402)	(0.431)	(0.521)	(0.537)
Old-age dependency ratio	0.0145**	0.0120*	-0.0187***	-0.0230***	0.0103	0.0114
	(0.00532)	(0.00553)	(0.00484)	(0.00526)	(0.00592)	(0.00625)
Female labor share	0.984***	0.965***	0.0754	0.283	0.763***	0.729***
	(0.167)	(0.184)	(0.164)	(0.187)	(0.180)	(0.190)
Population	170.6***	163.0***	136.8***	152.0***	182.2***	184.2***
	(35.55)	(33.46)	(25.75)	(31.66)	(41.49)	(42.10)
N	722	722	722	722	722	722
adj. R-sq	0.538	0.544	0.41	0.344	0.357	0.251

Table 1.4: Regression table of regression (1.5)

Regression Table: Equation (1.5) across different macro industries

This section investigates the heterogeneity of the specialization trade-off across different sectors of the economy. Table 1.5 presents the results, where the aggregate specialization from equation (1.5) is replaced by specialization within six major industries simultaneously. Note that these results are quantitatively similar to those obtained when estimating the equation for each industry individually.

The estimates reveal that the trade-off is not uniform and is heavily concentrated in specific industries. The pattern is most pronounced for Manufacturing and Wholesale Trade. Specialization in Manufacturing is associated with significantly lower long-run growth (-1.086) but higher initial income (2.296). This effect is even stronger and more precisely estimated for Wholesale Trade, which shows the largest negative coefficient on growth (-2.356) and a strong positive association with 1950 income (2.992).

In contrast, other sectors show different or insignificant patterns. Specialization in Retail and Transportation was associated with higher initial income but had no statistically significant relationship with long-run growth. Interestingly, a higher concentration in Services in 1950 is linked to significantly lower initial income levels, while specialization in Agriculture shows no significant association with either outcome. These results strongly suggest that the aggregate specialization trade-off documented in the main text is primarily driven by concentration in goods-producing and distributing sectors.

The standard set of control variables, including initial income, high-skill share, and population, remain highly significant across both columns, and their coefficients are quantitatively similar to those in the main specification (Table 1.4). The relative size of the sectors n 1950 is shown in Table 1.3.

	1950-2020 Growth	1950 Income level
	(1)	(2)
Specialization in		
Manufacturing	-1.086**	2.296***
	(0.373)	(0.488)
Services	1.441	-5.386**
	(1.201)	(1.776)
Agriculture	-0.0653	0.35
	(0.191)	(0.241)
Transportation	0.397	2.071*
	(0.912)	(0.894)
Wholesale	-2.356***	2.992**
	(0.698)	(0.994)
Retail	-0.747	3.404***
	(0.764)	(0.64)
1950 measures:		
$\hat{g}$	-0.249**	0.0241
	(0.0836)	(0.11)
log income p.c.	-0.863***	
	(0.0345)	
High-skill labor share	1.380***	5.687***
	(0.411)	(0.399)
Old-age dependency ratio	0.0163**	-0.0228***
	(0.00542)	(0.00492)
Female labor share	1.137***	-0.0526
	(0.186)	(0.194)
Population	165.7***	149.6***
N	722	722
adj. R-sq	0.54	0.399

Table 1.5: Regression table of regression (1.5) across different industries

Regression Table: Equation (1.5) with the Herfindahl index

Table 1.6 shows the regression results using the Herfindahl-Hirschman Index (HHI) as a measure of specialization. The results are in line with the baseline results: a greater specialization in 1950 is related to higher per capita income in 1950, but lower growth between 1950-2020. While both the HHI and the Gini coefficient measure concentration, they differ in their construction and sensitivity. The HHI is calculated as the sum of the squared income shares of all industries in a region. This method gives disproportionately more weight to the largest industries, making it particularly sensitive to the presence of one or two dominant sectors. In contrast, the Gini coefficient measures the inequality across the entire distribution of industry shares and is less influenced by extreme values. It provides a broader assessment of how evenly distributed economic activity is.

	1950-2020 Growth	1950 Income level
	(1)	(2)
Specialization	-0.212*	0.933***
	(0.122)	(0.134)
$\hat{g}$	-0.169***	-0.233**
	(0.0470)	(0.064)
log income	-0.886***	
	(0.0328)	
High-skill share	1.482***	6.351***
	(0.424)	(0.436)
Old-age dependency	0.0149**	-0.0213***
	(0.00532)	(0.00512)
Female share	1.04***	-0.172
	(0.167)	(0.179)
Population	172.7***	136.7***
	(35.77)	(25.67)
N	722	722
adj. R-sq	0.535	0.375

Table 1.6: Regression table of regression (1.5) with HHI

Regression Table: Equation (1.5) at different horizons

This section explores the temporal dynamics of the specialization trade-off by running a series of dynamic panel regressions. Instead of a single cross-section, here I regress future income per capita growth over a specific horizon (k years) on the level of specialization at the beginning of that period (in year $t-h$). The coefficient of interest β^h is estimated from the following regression:

$$y_{c,t} - y_{c,t-h} = \alpha^h + \beta^h \cdot Gini_{c,t-h} + \delta \cdot \hat{g}_{c,t-h} + \gamma' \cdot Z_{c,t-h} + \epsilon_{c,t-h} \tag{1.29}$$

where all controls correspond to the controls in equation (1.5) at horizon $t-h$. Table 1.7 reports the effect of specialization on income growth over different time horizons. These findings provide two crucial insights. First, the adverse effects of tradable specialization are not immediate but emerge and accumulate over decades, consistent with a story of slowing adaptation or declining dynamism in economies heavily reliant on a narrow set of tradable industries. The table shows that the adverse effect of specialization on

growth appears after a 20-year horizon. Second, the negative long-run consequence of specialization is driven specifically by concentration in tradable industries.

	Income pc growth						
	10-year (1)	20-year (2)	30-year (3)	40-year (4)	50-year (5)	60-year (6)	70-year (7)
Trad. Specialization (t-10)	0.00152						
	(-0.025)						
Trad. Specialization (t-20)		-0.0747***					
		(-0.0209)					
Trad. Specialization (t-30)			-0.106***				
			(-0.0254)				
Trad. Specialization (t-40)				-0.162***			
				(-0.0293)			
Trad. Specialization (t-50)					-0.1420**		
					(-0.0147)		
Trad. Specialization (t-60)						-0.152**	
						(-0.0347)	
Trad. Specialization (t-70)							-0.154***
							(-0.04)
N	3528	3563	2842	2123	1403	1007	700
adj. R-sq	0.101	0.113	0.152	0.219	0.308	0.403	0.549

Table 1.7: Regression table: Specialization on growth at different horizons

Map of exposure to structural change

Figure 1.15 displays the geographical distribution of predicted income growth, $\hat{g}_c$, across all U.S. commuting zones (CZs). As defined in equation (1.2), this variable captures the long-run growth (1950-2020) that a CZ was structurally exposed to, based purely on its initial 1950 composition of 1-digit industries and the subsequent national growth rates of those industries. The map reveals stark regional patterns that reflect the major structural transformations in the U.S. economy over the latter half of the 20th century. Regions with high predicted growth (indicated by darker shades) are heavily concentrated in the Sun Belt and the West Coast. These areas had a larger initial employment share in sectors that experienced high national growth, such as services and, later, technology. In contrast, regions with low or even negative predicted growth (lighter shades) are prominent throughout the Rust Belt, the industrial Midwest, and parts of the Great Plains. These CZs were historically specialized in manufacturing and agriculture, sectors that experienced significant relative decline or slower growth at the national level. The map thus provides a clear visual representation of the structural economic headwinds and tailwinds that different regions faced, underscoring the necessity of including $\hat{g}_c$ as a control in equation (1.5) to isolate the distinct economic effects of industrial specialization.

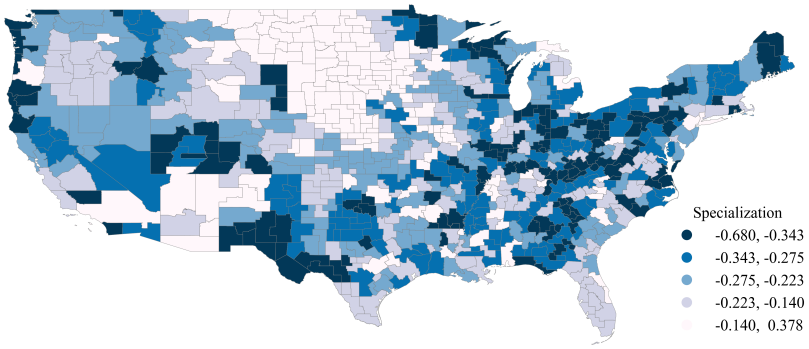


Figure 1.15: Commuting Zones by 1950 exposure to structural change

Notes: This figure shows a map of commuting zones binned by their 1950 estimate of $\hat{g}$ from equation (1.2).

Specialization trade-off at 20-year horizon

Figure 1.16 plots coefficients from two separate panel regressions estimated at a 20-year horizon. I show the 20-year horizon as the specialization trade-off appears at that horizon as shown above. Instead of a single linear coefficient, the effect of specialization is captured using dummy variables for 20 quantiles of the specialization Gini. The coefficients are estimated from the following regression:

$$y_{c,t-20} = \sum_{k=2}^{20} [\beta_k \cdot D_{k,c,t-20} + \delta \cdot \hat{g}_{k,c,t-20} + \gamma' \cdot Z_{k,c,t-20} + \epsilon_{k,c,t}] \tag{1.30}$$

where $D_{k,c,t-20}$ is a binary dummy equal to one if a commuting zone falls into the k-th quantile of specialization at time $t-20$. The figure illustrates the specialization trade-off: The left panel shows an upward-sloping relationship, indicating that commuting zones in higher quantiles of specialization have systematically higher initial income per capita. The right panel shows a downward-sloping relationship, meaning regions in higher specialization quantiles experience significantly lower income growth over the subsequent 20 years. This confirms that the benefits of specialization in terms of higher income are contemporaneous, while the costs in terms of lower growth emerge over longer horizons.

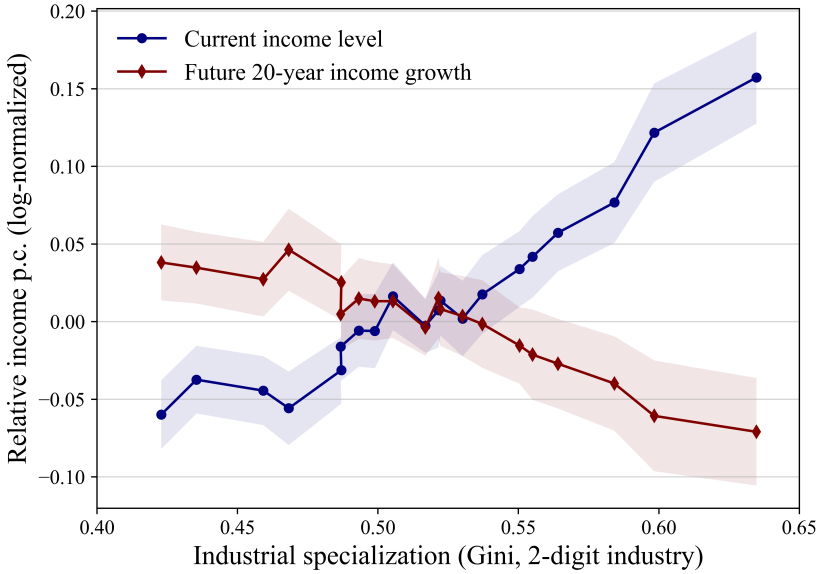


Figure 1.16: Industrial specialization, relative income and long-run growth

Notes: This figure plots the coefficients of regressions of 10-year commuting zone-level income growth and initial per capita income on dummies for 20 quantiles of the initial specialization, for the period 1950-2020. Specialization is computed as the Gini coefficient on 3-digit level industry income shares. 95% confidence intervals are shaded.

Map of regional specialization

This section provides a visual overview of how the geography of regional specialization has evolved in the United States over the last 70 years. The two figures map the concentration of economic activity across U.S. commuting zones at the beginning and end of the analysis period, grouping regions into five quintiles based on their specialization Gini coefficient, where darker shades represent a higher degree of industrial specialization.

The 1950 map shows that high levels of specialization were concentrated in the nation's industrial heartland. The darkest shades, indicating the most specialized regions, are located almost exclusively in the Rust Belt and the industrial Midwest, along with some areas in the Southeast. This pattern reflects an economy where regional dominance was defined by heavy manufacturing and specific industries, such as automobile production in and around Detroit, steel manufacturing in Pennsylvania, and textiles in the Carolinas.

By 2020, the economic landscape has transformed entirely. The once-specialized Rust Belt now exhibits significantly lower levels of industrial concentration. Instead, the new centers of high specialization are found predominantly on the East and West Coasts. This modern pattern is driven by knowledge-intensive and high-skill service industries. For instance, the highest levels of specialization are now seen in the technology sector of the San Francisco Bay Area, the financial and business services of the New York metropolitan area, and the entertainment industry of Los Angeles. Together, the two maps visually narrate the profound economic shift from a manufacturing-based economy to a knowledge- and service-based one.

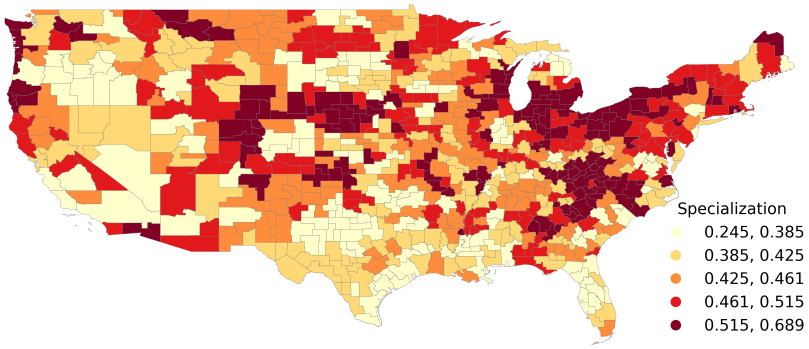


Figure 1.17: Commuting Zones by 1950 specialization

Notes: This figure shows a map of commuting zones binned by 1950 regional specialization as measured by the Gini coefficient on 1950 income shares by 3-digit industry.

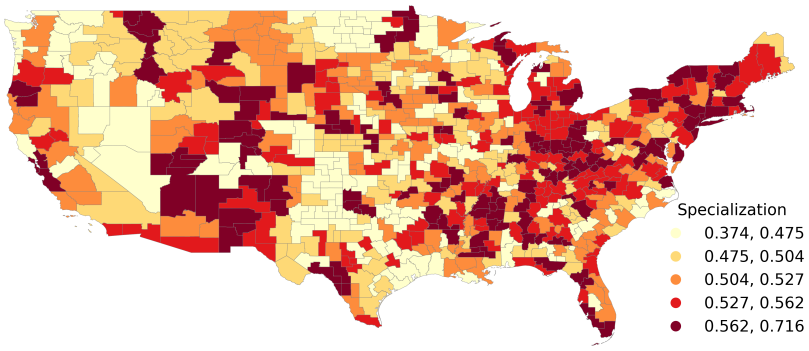


Figure 1.18: Commuting Zones by 2020 specialization

Notes: This figure shows a map of commuting zones binned by 2020 regional specialization as measured by the Gini coefficient on 2020 income shares by 3-digit industry.

Convergence and regional specialization

This section documents the evolution of regional income inequality in the U.S. since 1950, analyzing data at both the state and commuting zone levels. The analysis combines U.S. Census data for commuting zones with state-level data from the Bureau of Economic Analysis and housing value data from the NHGIS. Spatial dispersion is measured using the coefficient of variation of earnings per worker. To account for differences in local purchasing power, nominal income at the commuting zone level is deflated using a local housing price index, which serves as a proxy for the local price level.

To adjust for local price levels, I construct a housing price index for each commuting zone (CZ) and year using binned housing value data from the NHGIS. The process involves three steps. First, for each housing value bin provided in the data, I calculate a midpoint value. Second, for each CZ in a given year, I compute the mean housing value by taking a weighted average of these midpoints, where the weights are the number of

housing units in each respective bin. Finally, the housing price index is created by taking the logarithm of this mean value. This creates a consistent price index across time and space.

Figure 1.19 plots this measure of dispersion over time, illustrating a well-documented two-phase pattern. From 1950 until the early 1990s, the U.S. experienced a period of strong regional convergence, where the income gap between poorer and richer regions steadily narrowed, a finding consistent with the literature established by authors such as Barro and Sala-i-Martin (1992). However, since the 1990s, this trend has not only stalled but has reversed, leading to a new era of regional divergence where spatial income dispersion has increased, in some cases reaching levels higher than in the initial period.

This recent divergence can be driven by two distinct dynamics. One possibility is a persistent income hierarchy, where initially rich regions simply began growing faster than their poorer counterparts, thus increasing the distance between them without significant changes in their relative income rankings. An alternative explanation involves rank reversals, where the fortunes of regions have shifted over time, with some once-prosperous areas declining while other, previously poorer regions have risen. This is illustrated by the divergent paths of Dayton, Ohio, and Rochester, Minnesota, in Figure 1.1.

The theory developed in this paper offers a mechanism that can explain aspects of the latter story. The model captures how formerly rich regions, characterized by a high degree of specialization in specific industries, can experience long-term decline when those industries falter. This lock-in effect contrasts with the trajectory of other regions that, despite potentially lower initial specialization, possessed a more diverse industrial portfolio that facilitated greater adaptability and long-run growth.

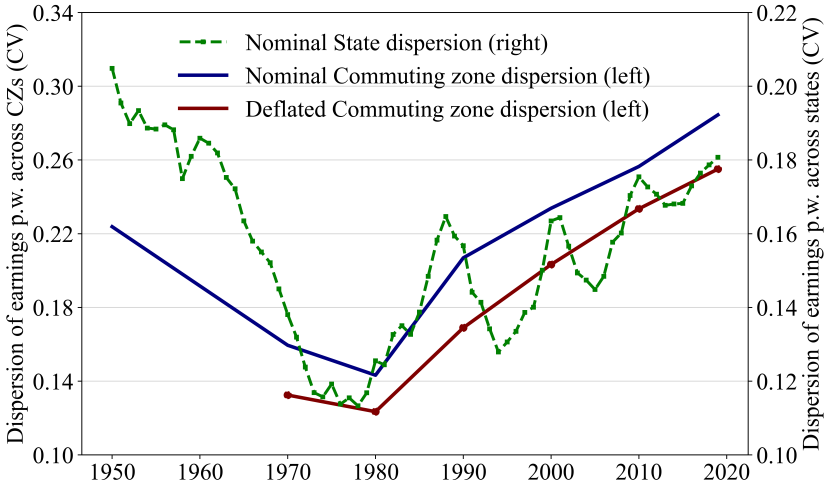


Figure 1.19: Dispersion of income p.c. across labor markets

Notes: This figure shows a spatial dispersion of per capita income across the U.S. since 1950.

Fact 2: Robustness and Extensions

In this section, I provide a range of robustness and extension exercises on Fact 2 as described in section 3.2.

Parametric specification

This section provides a parametric comparison to assess whether a standard linear model can adequately capture the relationship between regional specialization and income.

For this purpose, I perform a simple Ordinary Least Squares (OLS) regression of log-normalized specialization on relative per capita income using the following specification:

$$Gini_{c,t} = \alpha + \beta y_{c,t} + \epsilon_{c,t}$$

Figure 1.20 plots the linear fit derived from this regression against relative income, along with the 95 percent confidence interval. As is evident from the figure, the OLS model imposes a constant linear relationship between the two variables. This simple functional form is unable to capture the non-linear, S-shaped pattern identified in the main analysis, thereby motivating the use of more flexible non-parametric methods to characterize this relationship accurately.

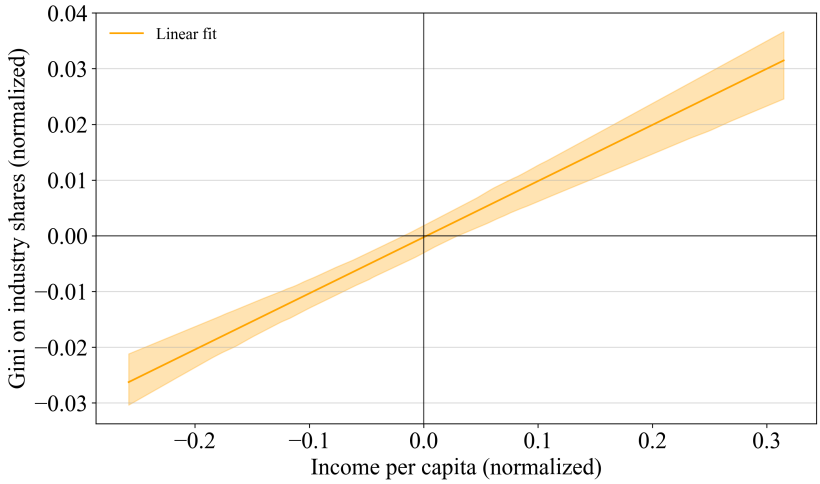


Figure 1.20: OLS estimate: Specialization vs. per capita income
Notes: The figure plots fitted values of normalized relative specialization against relative income. 95% bootstrapped confidence interval shown as a shaded area.

Specialization change on Change

This section provides a dynamic test of the relationship between specialization and income by examining how changes in one are related to changes in the other over time. To analyze this relationship in first differences, I estimate a 'change-on-change' regression for various horizons (h) using the following specification:

$$\log Gini_{c,t} - \log Gini_{c,t-h} = \alpha + \beta^h(\log y_{c,t} - \log y_{c,t-h}) + \gamma^h \log y_{c,t-h} + \epsilon_{c,t} \quad (1.31)$$

Here, the dependent variable is the log change in specialization over an h-year period. The main independent variable is the log change in per capita income over the same period, and the model also controls for the initial income level to account for any convergence effects.

Table 1.8 reports the estimates for the coefficient on the change in income, β^h. The key result is that the coefficient is positive and statistically significant across all time horizons, from 10 to 60 years. This indicates that regions experiencing faster income growth are also, on average, the ones whose economies become more specialized.

This finding provides strong dynamic support for Fact 2. The S-shaped relationship described in the main text implies that as regions develop and their incomes rise, their specialization levels also tend to increase. This regression confirms that very pattern by showing that periods of growth coincide with periods of rising specialization, demonstrating that the cross-sectional relationship holds dynamically within regions over time.

	Change in Specialization					
	10-year (1)	20-year (2)	30-year (3)	40-year (4)	50-year (5)	60-year (6)
Change in income pc.c						
(t-10)	0.313***					
	(0.0255)					
(t-20)		0.159***				
		(0.0321)				
(t-30)			0.264***			
			(0.0385)			
(t-40)				0.545***		
				(0.0422)		
(t-50)					0.512**	
					(0.0571)	
(t-60)						0393**
						(0.0487)
N	4332	3610	2888	2166	1444	722
adj. R-sq	0.064	0.047	0.087	0.275	0.327	0.113

Table 1.8: Regression table: Specialization change on income change

Fact 3: Robustness and Extensions

In this section, I provide a range of robustness and extension exercises on Fact 3 as described in section 3.2.

Persistence of Specialization at the regional level

This section evaluates the persistence of regional specialization over time using a rank-rank regression approach. For each period, commuting zones are ranked based on their specialization level, as measured by the log Gini coefficient of 3-digit industry income shares. The rank of a commuting zone at time t is then regressed on its rank at time $t-h$ for various horizons h, according to the following specification:

$$\text{Rank}_{c,t} = \alpha^h + \beta^h \cdot \text{Rank}_{c,t-h} + \epsilon_{c,t} \tag{1.32}$$

The coefficient of interest, β^h, is the rank-rank elasticity, which measures the degree of persistence. A value of one indicates perfect persistence, while a value of zero implies no relationship between past and future rankings.

Figure 1.21 plots the estimated elasticity β^h for horizons up to 70 years. The results show that regional specialization is highly persistent in the short run, but this persistence decays steadily over time. For instance, the estimate for a 60-year horizon is approximately 0.2. This implies that a 10 percent higher specialization rank is associated with only a 2 percent higher rank 60 years later. This demonstrates that while specialization rankings are sticky in the short to medium term, there is considerable mobility in the regional specialization hierarchy over the long run.

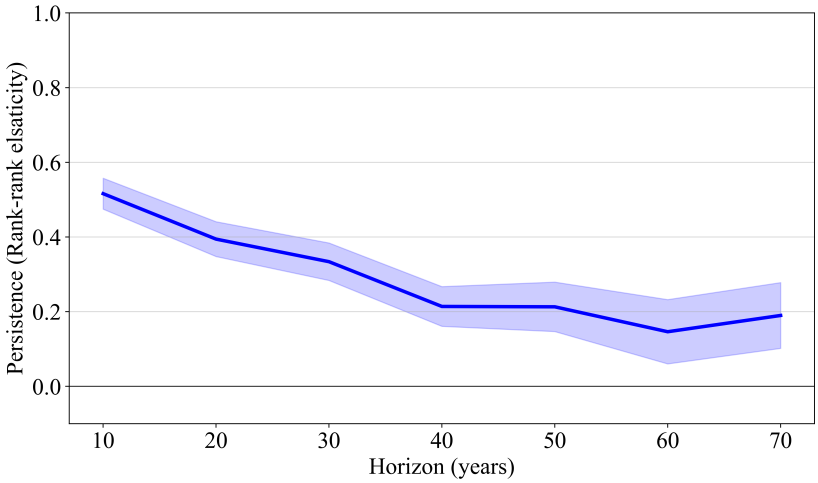


Figure 1.21: Persistence at the regional level

Notes: Figure 1.21 shows the rank-rank elasticity of regional specialization over a 70-year horizon, estimated from the regression in equation (1.5). The shaded area represents the 90% confidence interval.

Persistence of comparative advantages over time

The persistence of revealed comparative advantages (RCA) can be computed using the location-quotient - the ranking of sectors comparative advantage in a state over time. The revealed comparative advantage for each sector i in state c is given by $RCA_{i,c} = \left(\frac{X_{c,i}}{\sum_{i \in I} X_{c,i}} / \frac{X_{US,i}}{\sum_{i \in I} X_{US,i}}\right)$. Second, one can rank $RCA_{i,c}$ for every state-year and auto-regress it over different horizons as follows:

$$Rank_{c,i,t} = \alpha + \rho_{c,h} Rank_{c,i,t-h} + \epsilon_{c,i,t}$$

The following table shows the regression results. The ranking is highly persistence over short, medium and long-run horizons.

Horizon	1	5	10	20
Persistence	0.961***	0.8962***	0.839***	0.8148***
	(0.002)	(0.003)	(0.004)	(0.025)

Table 1.9: RCA persistence

Calibration: Further details

Agglomeration by industry

Table 1.10 reports the parameters for industry-specific agglomeration economies, ξ_i, which are used to calibrate the model in the main text. The estimates are taken directly from the study by Bartelme et al. (2019).

In this context, the parameter ξ_i represents the elasticity of an industry's productivity with respect to its own local size or density. For example, the value of 0.16 for the Food, Beverages and Tobacco industry implies that a 10 percent increase in the local scale of this sector is associated with a 1.6 percent increase in its productivity. These parameters are intended to capture the positive technological externalities that arise when firms in the same industry cluster together.

The original authors estimate these sector-level economies of scale using annual international trade data from 1995 to 2010 for the 15 2-digit manufacturing industries shown.

The estimates exhibit considerable variation across sectors, ranging from a low of 0.07 for Coke and Petroleum Products to a high of 0.25 for Rubber and Plastics, highlighting that the strength of agglomeration forces is highly industry-specific.

Industry	ξ_i
Food, Beverages and Tobacco	0.16
Textiles	0.12
Wood Products	0.11
Paper Products	0.11
Coke/ Petroleum Products	0.07
Chemicals	0.2
Rubber and Plastics	0.25
Mineral Products	0.13
Basic Metals	0.11
Fabricated Metals	0.13
Machinery and Equipment	0.13
Computers and Electronics	0.09
Electrical Machinery	0.09
Motor Vehicles	0.15
Other Transport Equipment	0.16

Table 1.10: Agglomeration parameter by industry

Capital adjustment costs by industry

This section details the parameters for industry-specific capital adjustment costs, Φ_i, used in the main analysis. Table 1.11 reports estimates from two empirical studies, Groth and Khan (2010) and Hall (2004).

These parameters govern the magnitude of convex costs that firms incur when they change their capital stock. The specific functional form for these costs used in this paper is:

$$\Phi_i(k_{i,t}, k_{i,t+1}) = \frac{\Phi_i}{2}\Big(\frac{k_{i,t+1}}{k_{i,t}} - 1 - \delta_i\Big)^2 k_{i,t} \quad \Phi_i \geq 0 \ \forall i \tag{1.33}$$

with depreciation rate δ_i. In this specification, a larger value of Φ_i implies that it is more costly for firms in industry i to rapidly adjust their capital stock.

The parameters from the cited studies were originally estimated from the first-order conditions of firm investment problems, using annual U.S. industry and firm-level data from 1948 to 2001. The parameters for capital adjustment costs reported in Table 1.11 reveal two key patterns with important implications for the model. First, the cost of adjusting capital is highly heterogeneous across sectors. The estimates vary substantially from one industry to another, and the two cited studies sometimes provide different rankings. For instance, the parameter for fabricated metals is 50% higher than for wooden products. This wide variation underscores the importance of using industry-specific parameters in a multi-sector model rather than a single, aggregate value.

Second, for many industries, the estimated parameters are small and often statistically insignificant from zero. This suggests that for a significant portion of the manufacturing sector, the physical costs of investment play a relatively minor role. This finding is consistent with the broader empirical literature, which often documents small estimated

adjustment costs (see, e.g., Hall (2004)). For the scoep of this paper, this implies that these physical frictions are likely not the primary driver of the strong persistence of specialization observed in the data. Other economic mechanisms, such as financial frictions or agglomeration economies, must therefore be at play.

Industry	Groth & Khan (2007)	Hall (2004)
Food, Beverages and Tobacco	0.01	-0.43
Textiles	3.26	-0.58
Wood Products	0.26	-0.22
Paper Products	0	-0.15
Coke/ Petroleum Products	-0.04	0.26
Chemicals	-0.08	-0.02
Rubber and Plastics	-0.2	-1.75
Mineral Products	0.1	0
Basic Metals	0.15	-0.06
Fabricated Metals	0.37	0.27
Machinery and Equipment	0.22	-0.92
Computers and Electronics	-0.01	0.38
Electrical Machinery	-0.01	-0.09
Motor Vehicles	0.25	0.4
Other Transport Equipment	-0.06	3.02

Table 1.11: Adjustment cost parameter by industry

1.8 Appendix - Model

Cross-derivatives in Planner optimality conditions

In recursive form, the cross-derivative of the bond decision and capital allocation decision rules are given by:

$$\Omega_i^B = \frac{1}{R}\Big[\frac{\partial u'(\mathcal{C}(b',\mathcal{K}',\mathcal{Z}'))}{\partial b'}(\mathcal{Q}(b',\mathcal{K}',\mathcal{Z}') + \alpha z_i' k_i'^{\xi_i+\alpha-1} - \phi_{1i}) + u'(\mathcal{C}(b',\mathcal{K}',\mathcal{Z}'))\Big(\frac{\partial \mathcal{Q}(b',\mathcal{K}',\mathcal{Z}')}{\partial b'} - \frac{\partial \phi_{1i}}{\partial b'}\Big) + \theta\Big(\frac{\partial q'}{\partial b'}\eta' - \frac{\partial \eta'}{\partial b'}\mathcal{Q}(b',\mathcal{K}',\mathcal{Z}')\Big)\Big] \tag{1.34}$$

$$\Omega_i^K = -u'(c)\phi_{22,i} + \frac{1}{R}\Big[\frac{\partial u'(\mathcal{C}(b',\mathcal{K}',\mathcal{Z}'))}{\partial k_i'}(\mathcal{Q}(b',\mathcal{K}',\mathcal{Z}') + \alpha z_i' k_i'^{\xi_i+\alpha-1} - \phi_{1i}) + u'(\mathcal{C}(b',\mathcal{K}',\mathcal{Z}'))\Big(\frac{\partial \mathcal{Q}(b',\mathcal{K}',\mathcal{Z}')}{\partial k_i'} + (\xi_i+\alpha-1)\alpha z_i' k_i'^{\xi_i+\alpha-2} - \phi_{11,i} - \phi_{12,i}\frac{\partial k_i''}{\partial k_i'}\Big) + \theta\Big(\frac{\partial q'}{\partial k_i'}\eta' - \frac{\partial \eta'}{\partial k_i'}\mathcal{Q}(b',\mathcal{K}',\mathcal{Z}')\Big)\Big] \tag{1.35}$$

Equation (1.34) captures the effects of a change in bond holdings today on the current and future price of capital through the cross-derivatives of the implementability constraints. It consists of three components. The first, captures how an extra unit of b_{t+1} affects future consumption and thus the discounting of future asset returns through the marginal utility. The second includes the effects by which b_{t+1} alters the marginal return of capital (i.e. dividends) in industry i. The third shows how it affects the future tightness of the collateral constraint.

Condition (1.35) captures the effects of a change in capital allocated to industry i on the current and future price of capital through the cross-derivatives of the implementability constraints. It consists of four terms. The first is the curvature on the adjustment costs $\phi_{22,i}$. The second, captures the effect of an increase in $k_{i,t+1}$ on future consumption, thus affecting the discounting of future asset returns. The third captures the effect by which a higher $k_{i,t+1}$ alters future dividends. the consists both of the curvature of the production function as well as the agglomeration externality ξ_i. The final terms shows how an additional unit of capital loosens the future collateral constraint.

Proof: Proposition 1

Proposition 1 (Decentralization with taxes and subsidies)**.** *The constrained-efficient equilibrium can be decentralized with a state-contingent tax on debt, and industry-specific tax and subsidies with tax revenue rebated as a lump-sum transfer and the tax rates set to satisfy:*

$$1+\tau_t^B = \frac{1}{\mathbb{E}_t[u'(t+1)]}\Big[\sum_i \delta_{i,t}\Omega_{i,t+1}^B - \frac{u''(t+1)}{u'(t+1)}\theta\eta_{t+1}q_{t+1} - u''(t+1)\sum_i \delta_{i,t+1}\phi_{i,t+1}^2\Big] + \frac{1}{\beta R_t \mathbb{E}_t[u'(t+1)]}\mathbb{E}_t\Big[\frac{u''(t)}{u'(t)}\theta\eta_t q_t + u''(t)\sum_i \delta_{i,t}\phi_{i,t}^2\Big] \tag{1.36}$$

$$1+\tau_{i,t}^K = \frac{1}{\mathbb{E}_t[\alpha z_{i,t+1}k_{i,t+1}^{\alpha+\xi_i-1}]}\Big[\Big(u'(t+1) - \frac{u''(t+1)}{u'(t+1)}\theta\eta_{t+1}q_{t+1} - u''(t+1)\sum_i \delta_{i,t+1}\phi_{i,t+1}\Big)((\alpha+\xi_i)z_{i,t+1}k_{i,t+1}^{\alpha+\xi_i-1} - \phi_{i,t+1}^1) + \sum_i \delta_{i,t}\Omega_{i,t+1}^K$$

$$- u'(t)(q_{t+1} - \phi^1_{i,t}) - \theta\eta_{t+1}q_{t+1}\Big]$$
$$+ \frac{1}{\beta\mathbb{E}_t[\alpha z_{i,t+1}k^{\alpha+\xi-1}_{i,t+1}]}\Big(u'(t)q_t + \Xi + \phi^2_{i,t}u''(t)\sum_i \delta_{i,t}\phi^2_{i,t}\Big) \tag{1.37}$$

where the arguments of the functions have been shorthanded as dates to keep the expression simple.

Define the taxes as:

$$1 + \tau^B_t = \frac{1}{\mathbb{E}_t[u'(t+1)]}\Big[\sum_i \delta_{i,t}\Omega^B_{i,t+1} - \frac{u''(t+1)}{u'(t+1)}\theta\eta_{t+1}q_{t+1} - u''(t+1)\sum_i \delta_{i,t+1}\phi^2_{i,t+1}\Big]$$
$$+ \frac{1}{\beta R_t\mathbb{E}_t[u'(t+1)]}\mathbb{E}_t\Big[\frac{u''(t)}{u'(t)}\theta\eta_t q_t + u''(t)\sum_i \delta_{i,t}\phi^2_{i,t}\Big] \tag{1.38}$$

$$1 + \tau^K_{i,t} = \frac{1}{\mathbb{E}_t[\alpha z_{i,t+1}k^{\alpha+\xi_i-1}_{i,t+1}]}\Big[\Big(u'(t+1) - \frac{u''(t+1)}{u'(t+1)}\theta\eta_{t+1}q_{t+1} - u''(t+1)\sum_i \delta_{i,t+1}\phi_{i,t+1}\Big)$$
$$((\alpha+\xi_i)z_{i,t+1}k^{\alpha+\xi_i-1}_{i,t+1} - \phi^1_{i,t+1}) + \sum_i \delta_{i,t}\Omega^K_{i,t+1}$$
$$- u'(t)(q_{t+1} - \phi^1_{i,t}) - \theta\eta_{t+1}q_{t+1}\Big]$$
$$+ \frac{1}{\beta\mathbb{E}_t[\alpha z_{i,t+1}k^{\alpha+\xi-1}_{i,t+1}]}\Big(u'(t)q_t + \Xi + \phi^2_{i,t}u''(t)\sum_i \delta_{i,t}\phi^2_{i,t}\Big) \tag{1.39}$$

I prove the proposition by showing that the decentralized equilibrium with the taxes yields the same optimality conditions as the planner's constrained-efficient equilibrium. The constrained efficient equilibrium can be characterized by sequences $\{c_t, k_{t+1}, b_{t+1}, q_t, \lambda^*_t, \eta_t\}^\infty_{t=0}$ that satisfy equations (1.11), (1.12) (1.13), (1.20), (1.21), (1.23), (1.22), $k_t = 1$ together with the complementary slackness conditions. The regulated decentralized equilibrium is characterized by sequences $\{c_t, k_{t+1}, b_{t+1}, q_t, \lambda_t\eta_t\}^\infty_{t=0}$ that satisfy equations (1.10), (1.11), (1.12) (1.13), $k_t = 1$ together with the complementary slackness conditions. Substitution the expression for the debt tax (1.38) and (1.25) yields condition (1.23) and identical conditions characterizing the two equilibria. Likewise, substituting the expression for the capital tax (1.39) and (1.24) yields condition (1.22) and identical conditions characterizing the two equilibria.

Solution algorithm

Following Bianchi and Mendoza (2018), I use a time iteration algorithm, modified to address the occasionally binding endogenous constraint and the additional number of state variables due to the multi-industry setting. Formally, the computation of the competitive equilibrium requires solving for functions
$\{\mathcal{B}(b,\mathcal{K},\mathcal{Z}), \mathcal{Q}(b,\mathcal{K},\mathcal{Z}), \mathcal{C}(b,\mathcal{K},\mathcal{Z}), \mathcal{K}_i(b,\mathcal{K},\mathcal{Z}), \eta(b,\mathcal{K},\mathcal{Z})$ such that:

$$\mathcal{C}(b,\mathcal{K},\mathcal{Z}) + \frac{\mathcal{B}(b,\mathcal{K},\mathcal{Z})}{R} = \sum_i (z_i f(\mathcal{K}_i(b,\mathcal{K},\mathcal{Z})) - \Phi_i(.)) + b \tag{1.40}$$

$$-\frac{\mathcal{B}(b,\mathcal{K},\mathcal{Z})}{R} \leq \theta\mathcal{Q}(b,\mathcal{K},\mathcal{Z}) \tag{1.41}$$

$$u'(\mathcal{C}(b,\mathcal{K},\mathcal{Z})) = \beta R\mathbb{E}[u'(\mathcal{C}(\mathcal{B}(b,\mathcal{K}_i(b,\mathcal{K},\mathcal{Z}),\mathcal{Z}')] + \eta(b,\mathcal{K},\mathcal{Z}) \tag{1.42}$$

$$u'(c)(q+\phi_i^2(.)) = \beta\mathbb{E}[u'(\mathcal{C}(\mathcal{B}(b,\mathcal{K}_i(b,\mathcal{K},\mathcal{Z}),\mathcal{Z}')(\mathcal{Q}(b,\mathcal{K},\mathcal{Z})+z_i'f'(\mathcal{K}_i(b,\mathcal{K},\mathcal{Z})) - \phi_i^1(.)) + \theta\mathcal{Q}(b,\mathcal{K},\mathcal{Z})\eta(b,\mathcal{K},\mathcal{Z})] \quad \forall i = \{1,...,I\} \tag{1.43}$$

The algorithm follows these steps:

1. Generate a discrete grid for the economy's bond position $G_b = \{b_1, ..., b_M\}$, the capital allocation $G_{k_i} = \{k_1, ..., k_M\}$ and the shock state space $G_{z_i} = \{z_1, ..., z_M\}$. in addition, generate a discrete grid for the economy's relative position of current levels of relative productivities $G_{\Lambda_i} = \{\Lambda_1, ..., \Lambda_M\}$. Choose an I use a piecewise linear function approximation and uniformly spaced grids.
2. Conjecture $\{\mathcal{B}_j(b,\mathcal{K},\mathcal{Z}), \mathcal{Q}_j(b,\mathcal{K},\mathcal{Z}), \mathcal{C}_j(b,\mathcal{K},\mathcal{Z}), \mathcal{K}_{i,j}(b,\mathcal{K},\mathcal{Z}), \eta_j(b,\mathcal{K},\mathcal{Z})$
3. Set $l = 1$.
4. Solve for the values of $\{\mathcal{B}_{j-l}(b,\mathcal{K},\mathcal{Z}), \mathcal{Q}_{j-l}(b,\mathcal{K},\mathcal{Z}), \mathcal{C}_{j-l}(b,\mathcal{K},\mathcal{Z}), \mathcal{K}_{i,j-l}(b,\mathcal{K},\mathcal{Z}), \eta_{j-l}(b,\mathcal{K},\mathcal{Z})$ at time $j-l$ functions above across all grids.
   (a) Solve for the capital allocation combing the capital Euler equations and the market clearing condition into no-arbitrage conditions and using a roof finding algorithm.
   (b) Assume the collateral constraint is not binding. Set $\eta_{j-l}(b,\mathcal{K},\mathcal{Z}) = 0$ and solve for $\mathcal{B}_j(b,\mathcal{K},\mathcal{Z})$ and $\mathcal{C}_j(b,\mathcal{K},\mathcal{Z})$ using the FOCs.
   (c) Check whether $-\frac{\mathcal{B}(b,\mathcal{K},\mathcal{Z})}{R} \leq \theta\mathcal{Q}(b,\mathcal{K},\mathcal{Z})$ holds using the asset price from the previous iteration to ensure stability.
   (d) If the constraint is satisfied, move to the next grid point.
   (e) Otherwise, solve for $\eta_{j-l}(b,\mathcal{K},\mathcal{Z})$ with equality.
   (f) Solve for $\mathcal{Q}_j(b,\mathcal{K},\mathcal{Z})$.
5. Evaluate convergence error and iterate until convergence.

Chapter 2

Monetary Policy in Currency Unions with Unequal Countries

Lukas Boehnert
University of Oxford

Sergio de Ferra
University of Oxford and CEPR

Kurt Mitman
IIES, CEMFI, CEPR, and IZA

Federica Romei
University of Oxford and CEPR

Email: lukas.boehnert@economics.ox.ac.uk, sergio.deferra@economics.ox.ac.uk, kurt.mitman@iies.su.se, federica.romei@economics.ox.ac.uk.

An earlier version of this paper was presented under the title "Inequality, Expenditure Composition and the Transmission of Monetary Policy". We thank for precious feedback and suggestions Christian Bayer, Florin Bilbiie, Davide Debortoli, Andrea Ferrero, Katja Mann, Dmitry Mukhin, Gianluca Violante, and seminar participants at Banco de España, Bonn, CREI, CEMFI, European Central Bank, ESSIM, Konstanz Seminar, Midwest Macro, Oxford, SED, Sheffield, UT Austin. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant agreement No. 759482) and the Swedish Research Council Project Grant 2023-01635. Mitman gratefully acknowledges the support of the Ragnar Söderbergs stiftelse, the Riksbankens Jubileumsfond, the Ministerio de Ciencia, Innovación y Universidades Grant IneqEx. Romei gratefully acknowledges financial support from the UK Research and Innovation joint European Research Council grant (CUR01080-HetAgHetCou).

Abstract

We investigate how the composition of expenditure shapes the transmission of monetary policy in a currency union. European Monetary Union data reveal three facts: (1) higher inequality countries have larger service expenditure shares; (2) monetary policy has a weaker output impact in these high-service-share, high-inequality countries; and (3) monetary policy induces systematic trade flows between high- and low-service-share countries. We develop a New Keynesian model with non-homothetic preferences and heterogeneous sectoral income that rationalizes these facts. Pro-cyclical inequality, driven by wealthier households' greater income exposure to services, buffers poorer households' consumption to contractionary shocks, dampening overall policy transmission. Our findings suggest that accounting for cross-country differences in consumption and income distributions is essential for understanding common monetary policy.

2.1 Introduction

The European Monetary Union (EMU) spans countries with different levels of per-capita income and inequality, presenting a significant challenge for common monetary policy. As the EMU has expanded, these cross-country heterogeneities have become more pronounced, raising a key question: How do differences in the level and distribution of income across member states affect the efficacy of a single monetary policy? This paper investigates this question, focusing on an expenditure composition channel. We argue that heterogeneity in income and its distribution systematically shapes how economies respond to monetary policy, primarily through its influence on the relative consumption of goods and services.

Our point of departure is to document empirical regularities related to the composition of consumption expenditures within and across countries in the EMU. We begin by confirming the findings of a large literature on structural transformation (e.g., Chenery, 1960; Clark, 1940; Kuznets, 1957): within EMU countries higher-income households allocate a greater proportion of their spending to services. Building on this foundation, we document three novel empirical regularities about the EMU.

First, at the aggregate level, countries with higher income inequality allocate a larger share of their total expenditure to services. This pattern is closely linked to our initial observation: high-income households, who account for a larger share of total income in more unequal countries, allocate a larger portion of their expenditure to luxuries (services), resulting in a higher aggregate expenditure share on services.

Second, these differences in expenditure composition have direct consequences for monetary policy transmission. Using local projections (Jorda, 2005), we show that countries with a higher average share of expenditure on services exhibit a *weaker* output response to identified monetary policy shocks. This finding is striking. The service sector typically exhibits greater price rigidity than the goods sector (see, e.g. Alvarez et al., 2006; Cravino et al., 2020; Gautier et al., 2024), which, all else equal, should amplify the real effects of monetary policy in economies with large service sectors. Yet, we find the contrary: a higher service expenditure share is associated with a *weaker* monetary policy transmission to output.

Third, monetary policy shocks induce heterogeneous responses in external balances. Using the same methodology as above, we document that countries with high service expenditure shares tend to run trade deficits following a contractionary shock, while those with low service shares run surpluses.

We rationalize these empirical findings through the lens of a multi-country New Keynesian model. Our framework incorporates: cross-country variation in income and inequality, non-homothetic preferences, where services are luxuries, incomplete financial markets, and unequal incidence of sectoral income, allowing wealthier households' income to be disproportionately exposed to the service sector. Our choice of non-homothetic preferences is a natural way (motivated by models in the structural change literature e.g., Boppart (2014); Herrendorf et al. (2014); Kongsamut et al. (2001)) to generate higher service expenditure shares in higher-income and more unequal countries, consistent with our first empirical regularity. To keep the model parsimonious, we consider only two types of households in the model economy: Ricardian and Hand-to-Mouth (HtM). The Ricardian household has access to financial markets, while the HtM household does not. Both earn income from goods and service sectors, but their relative income shares from these sectors can differ across countries, as observed empirically.

These differences in income distributions lead to differences in the cyclicality of relative household incomes, and specifically in the response of income inequality to aggregate shocks.

The model's ability to generate dampened monetary policy transmission observed in high service-share countries (our second empirical regularity) hinges on a mechanism of *pro-cyclical inequality*. This arises when wealthier households derive a larger share of their income from the service sector. Such an income structure not only generates higher inequality and larger service shares in those countries but also makes the earnings of Ricardian households more pro-cyclical than those of HtM households. These differences in income structure are key to generating both the cross-country variation in aggregate service shares (via inequality) and the heterogeneous cyclical responses in relative household incomes that drive pro-cyclical inequality and weaken the transmission of monetary policy.

We first analyze a homogeneous monetary union (akin to a closed economy). Here, comparative statics with the model demonstrate that higher inequality (driven by Ricardian exposure to services) indeed leads to both a higher aggregate service expenditure share and a weaker transmission of monetary policy to output. This provides a direct, joint rationalization for our first two empirical regularities.

Extending the model to a heterogeneous monetary union where countries differ in their income distributions (and thus in their baseline service shares), the model further predicts the emergence of systematic trade imbalances following a common monetary policy shock, consistent with our third empirical regularity. Unequal, high service-share countries experience a milder consumption contraction due to the pro-cyclical inequality mechanism and run trade deficits, while more equal, lower service-share countries run surpluses. This occurs even as goods output falls similarly across the union, highlighting how domestic distributional dynamics and expenditure composition spill over into external positions.

Literature Review. This paper contributes to the macroeconomic literature on the transmission of monetary policy with heterogeneous agents, and on non-homothetic preferences, consumption, and household heterogeneity.

We contribute to the literature on monetary policy with heterogeneous agents empirically and theoretically. Empirically, we document cross-country differences in monetary policy transmission and their systematic links to the composition of consumption. Theoretically, we introduce non-homothetic preferences and heterogeneity across countries in the distribution and composition of income, in a heterogeneous-agent, open-economy, New-Keyensian model. We show that the combination of non-homothetic preferences, market incompleteness, and income composition heterogeneity can generate pro-cyclical inequality, weakening monetary policy transmission relative to a representative-agent model. We thus contribute to the large and growing macroeconomics literature on heterogeneous agents and monetary policy, including seminal works by Kaplan et al. (2018), Auclert (2019), Gornemann et al. (2021) and Krueger et al. (2010). Our two-agent model allows for an analytical characterization of the transmission of monetary policy, thus expanding on the insights from the work by Debortoli and Galí (2024), Bilbiie (2024) and Acharya and Dogra (2020). Our open-economy results build on our previous work on the implications of international differences in inequality for countries' external balance, de Ferra et al. (2021), and on the international transmission of monetary policy in

heterogeneous-agent economies in de Ferra et al. (2020).[1]

Our contribution to the literature on non-homothetic preferences is again twofold. First, we document empirically differences in the composition of consumption across households and countries in the euro area, and we relate these systematic differences to cross-country differences in average income and in its distribution. Second, we introduce non-homothetic preferences in a model of a monetary union with heterogeneous agents, and we show how these preferences interact with income inequality and market incompleteness to shape the transmission of monetary policy. Our theoretical contribution builds on the seminal work by Deaton and Muellbauer (1980) and incorporates the description of preferences introduced by Boppart (2014) to characterize long-run trends in the composition of consumption between luxuries and necessities.

Several recent papers have analysed the implications of non-homothetic preferences in heterogeneous-agent models. Straub (2019) consider non-homothetic preferences in the context of saving rates, Cravino et al. (2020) study heterogeneity in the nominal rigidity of consumption baskets, Jaravel (2018) documents heterogeneity in inflation fluctuations experienced by different households in the income distribution and Jaimovich et al. (2019) show that consumption shifts towards lower-quality goods exacerbate recessions. Our paper contributes to this literature by showing that when households differ in terms of the composition of their income from different sources, non-homothetic preferences give rise to heterogeneous income fluctuations across households, and thus may weaken the transmission of monetary policy.

We also relate to the recent paper by Olivi et al. (2024), who study optimal monetary policy in a model with non-homothetic preferences and household heterogeneity. Relative to that paper, we show that heterogeneity across households in the severity of income fluctuations is a key driver of the aggregate implications of non-homothetic preferences for the transmission of monetary policy. Another recent contribution in the field is the work by Ferriere et al. (2024), who analyze the optimal conduct of fiscal policy in an economy with long-run changes in inequality and non-homothetic preferences.

2.2 Empirical Analysis

We document three empirical facts regarding expenditure composition, the income distribution, and monetary policy transmission in the EMU. These facts motivate and guide our theoretical analysis.

First, we document that countries with higher income inequality tend to allocate a larger share of total consumption to services. Second, we show that economies with lower service expenditure shares exhibit stronger output responses to monetary policy shocks. Third, we show that monetary policy shocks induce heterogeneous responses in external balances across countries in the EMU.

2.2.1 Data

We combine data on household consumption expenditure, income levels, macroeconomic aggregates, and monetary policy shocks for 19 European Monetary Union (EMU) member states.

[1] A growing literature studies the implications of household heterogeneity for the transmission of shocks in open economies. See, inter alia Bayer et al. (2024), Guo et al. (2023), Ferrante and Gornemann (2022), Auclert et al. (2021), Hong (2020), Oskolkov (2023), Guntin et al. (2023).

Household consumption expenditure across income levels is primarily from the quinquennial Household Budget Survey (HBS, Eurostat), covering 1999-2020, which provides expenditure shares on three-digit COICOP categories per income quintile.[2] We convert these quintiles to disposable income levels using Eurostat income distribution data.[3]

Consumption expenditure is classified as goods (G) and services (S) based on COICOP (see Appendix 2.6 for Table 2.4), excluding housing. Aggregate service consumption shares are calculated annually. Income inequality is measured by the Gini coefficient (World Inequality Database, WID). Quarterly macroeconomic data (real GDP, value added by NACE Rev.2, price levels, trade balances, etc.) are from Eurostat. Value added is also classified as goods/services (Appendix 2.6). Household-level wealth and income-by-industry data are from the Household Finance and Consumption Survey (HFCS), used to estimate Hand-to-Mouth (HtM) shares and sectoral income differences. Monetary policy shocks are high-frequency surprises from Jarociński and Karadi (2020). Our sample for local projections includes countries post-ERM II entry, following Dorrucci et al. (2022).[4]

Finally, we gather household-level data on wealth and income by industry from the Household Finance and Consumption Survey (HFCS). We use this data to compute the share of Hand-to-Mouth (HtM) households by country and to estimate differences in sectoral income shares between constrained and unconstrained households.

2.2.2 Empirical Regularities

A well-documented pattern, consistent with theories of structural transformation and non-homothetic preferences (Boppart, 2014; Herrendorf et al., 2014; Kongsamut et al., 2001), is that households with higher incomes allocate a systematically larger share of their consumption expenditure to services. We confirm this for our sample of euro-area countries. At the micro level, Figure 2.1 shows that the average service expenditure share rises from under 20% for the lowest income quintiles to nearly 50% for the highest. This underscores the relevance of non-homotheticity, with services functioning as luxury goods, as a foundational element for our subsequent analysis.

Disaggregating consumption into finer (2-digit COICOP) categories further underscores this non-homotheticity. Estimating the relationship between income and expenditure shares ($\omega_{t,i,j} = \alpha_i + \beta_i I_{t,j} + \phi_{i,j} + \psi_{i,t} + \epsilon_{t,i,j}$, controlling for country and time fixed effects), Figure 2.2 shows that the income coefficient (β_i) is significantly positive for most service categories (e.g., restaurants) and significantly negative for key tradable necessities like food.

Empirical Regularity 1: Services consumption increases with income inequality

At the aggregate level, countries with higher income inequality tend to have a larger expenditure share on services. We document this by estimating the following weighted

[2]Note that for Bulgaria, the Czech Republic, Estonia, Croatia, Cyprus, Latvia, Lithuania, Hungary, Malta, Poland, Romania, Slovenia and Slovakia the sample starts in 2005 due to their later accession to the European Union. For Italy, data exists only until 2010.

[3]Income statistics are given as top cut-off points for each decile. We match the income level for each quintile with the top cut-off point of the first corresponding decile (e.g., the second quintile, spanning 3rd and 4th deciles, is matched with the 4th decile cut-off; the fifth quintile with the 9th decile cut-off).

[4]As a result, we exclude Cyprus, Estonia, Latvia, Lithuania, Malta, and Slovenia before 2004, and Slovakia before 2005.

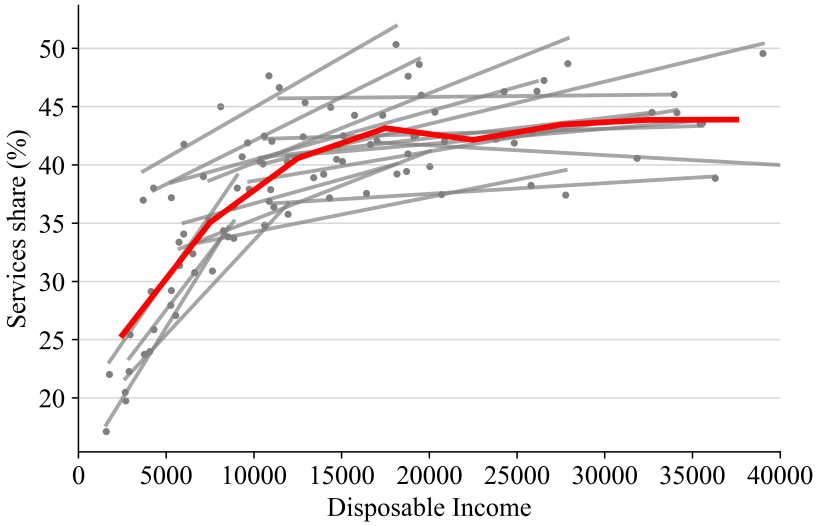


Figure 2.1: Services expenditure shares by household income level.

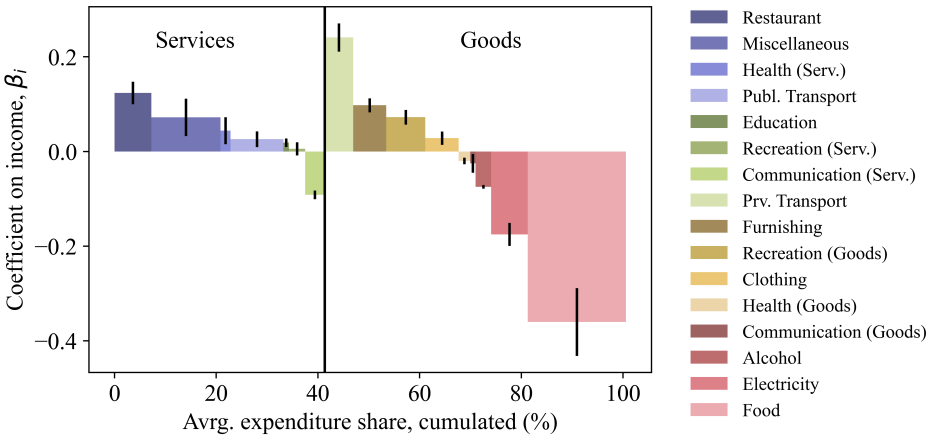


Figure 2.2: Income elasticity of consumption shares by COICOP category.

(by GDP) linear regression:

$$\bar{\omega}_{N,k} = \alpha + \beta \overline{Gini}_k + \gamma_i \bar{X}'_k + \epsilon_k, \tag{2.1}$$

where $\bar{\omega}_{N,k}$ is the average service consumption share, $\overline{Gini}_k$ is the Gini coefficient of income, and $\bar{X}$ is a vector of controls including the trade balance, demographics, government size, and GDP per capita. Figure 2.3 and Table 2.1 show a positive and significant relationship: greater inequality correlates with higher service shares. This is consistent with our findings that services are luxury goods if higher-income households, who spend more on services, account for a larger share of total expenditure in more unequal countries.

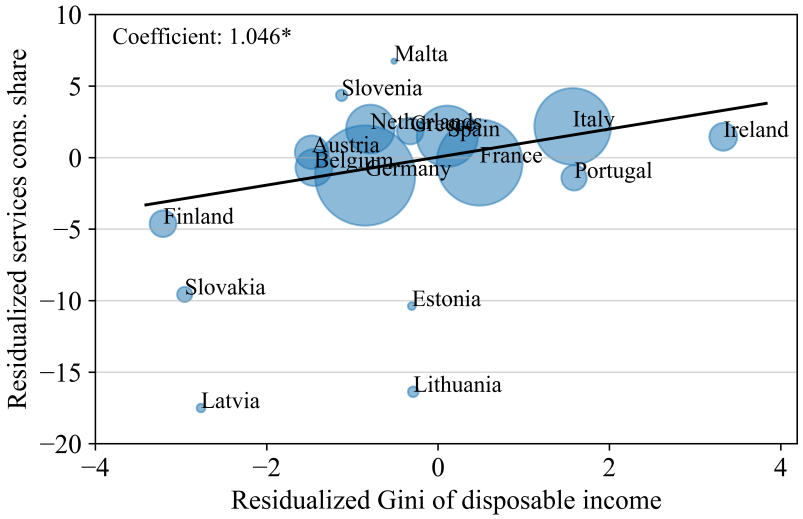


Figure 2.3: Inequality and Services Expenditure Shares.

	Services expenditure share		
Gini	1.046*		
	(-0.397)		
p90p10		6.128**	
		(-1.596)	
p75p25			13.36**
			(-3.91)
Trade balance	-1.034	-0.658	-0.769
	(-0.664)	(-0.529)	(-0.534)
Old-age dependency	-0.764**	-0.769***	-0.679**
	(-0.209)	(-0.173)	(-0.166)
Gov. exp.	-0.246	0.0512	-0.0482
	(-0.364)	(-0.288)	(-0.286)
GDP p.c.	0.648	0.625*	0.597*
	(-0.327)	(-0.231)	(-0.244)
N	17	17	17
adj. R-sq	0.571	0.689	0.644

Table 2.1: Coefficient Estimates from Regression (2.1).

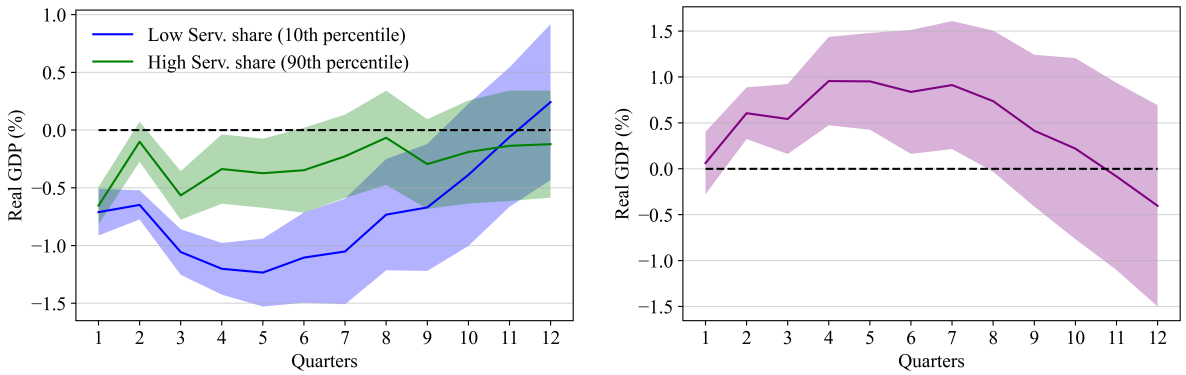
Empirical Regularity 2: Weaker monetary policy response in countries with high services consumption shares

Countries with a higher average share of expenditure on services exhibit a significantly weaker output response to monetary policy shocks. We establish this fact using local projections (Jorda, 2005; Montiel Olea and Plagborg-Møller, 2021) to estimate the impact of identified monetary policy shocks (i_t) on real GDP ($y_{t,k}$), interacted with the average service consumption share ($\bar{\omega}_k$):

$$y_{t+h,k} - y_{t-1,k} = \alpha + \beta^h i_t + \gamma^h (i_t \times \bar{\omega}_k) + \sum_{s=1}^{p} \Gamma_s^h y_{t-s,k} + \phi_k + u_{t+h,k}. \tag{2.2}$$

for each country k in the EMU. For inference, we follow the approach proposed by Montiel Olea and Plagborg-Møller (2021). In particular, we include $p = 3$ lags of the dependent variable as regressors to deal with potential auto-correlation and we cluster standard errors around countries. Finally, we also control for country fixed effects ϕ_k and an intercept α. Figure 2.4 shows that the output decline following a contractionary shock is substantially larger and more persistent in countries with low service shares (e.g., 1.2% cumulative decline after five quarters at the 10th percentile of service share) compared to those with high service shares (0.5% decline at the 90th percentile). The interaction term γ^h is negative and significant, confirming this differential response.

We provide an extensive discussion of robustness and extensions in Appendix 2.6. We also show in Appendix 2.6 that our impulse response functions for individual countries closely align with estimates found in the existing literature.



(a) High and low service share

(b) Difference (Interaction Term γ^h)

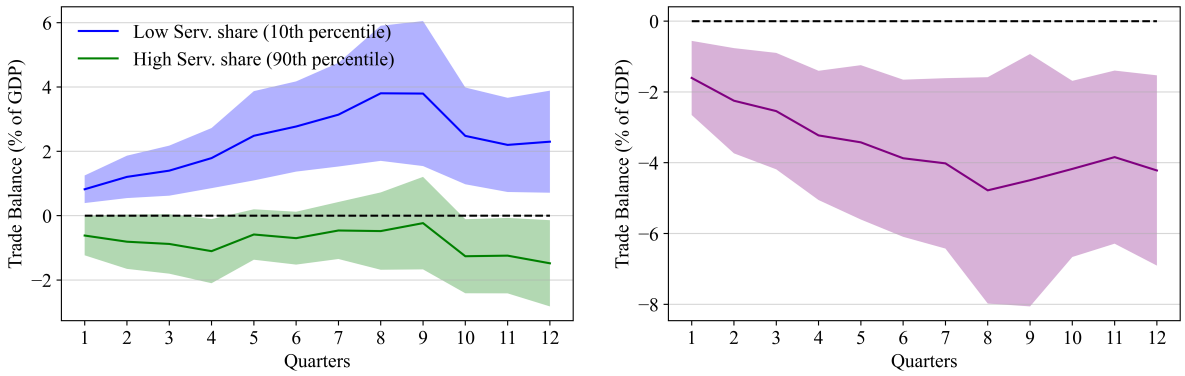
Figure 2.4: Impulse response of real GDP to a contractionary monetary policy shock.

Empirical Regularity 3: Monetary policy shocks induce heterogeneous trade balance responses

The expenditure composition across countries also correlates with how their trade balances respond to monetary policy shocks. We find that countries with relatively low service consumption shares experience a larger *improvement* (or smaller deterioration) in their trade balance following a contractionary monetary policy shock. To show this, we again use local projections, estimating the cumulative response of the trade balance (as a % of GDP) to monetary policy shocks, interacted with the service consumption share $\bar{\omega}_k$:

$$\sum_{j=0}^{h}(tb_{t+j,k} - tb_{t-1,k}) = \alpha^h + \beta^h \sum_{j=0}^{h} i_{t+j} + \gamma^h \left(\sum_{j=0}^{h} i_{t+j} \times \bar{\omega}_k \right) + \cdots + u_{t+h,k}. \tag{2.3}$$

Figure 2.5 illustrates this. Following a one standard deviation contractionary shock, countries at the 10th percentile of service share see their trade balance improve by about 0.5 percentage points of GDP after two quarters. In contrast, countries at the 90th percentile of service share experience a slight deterioration. The difference, captured by the interaction term γ^h, is positive and statistically significant.



(a) High and low service share

(b) Difference (Interaction Term)

Figure 2.5: Cumulative impulse response of trade balance (% of GDP) to a contractionary monetary policy shock.

Notes: Countries with low service consumption shares experience a larger increase in the trade balance. The shaded areas represent 90% confidence intervals. Standard errors are clustered around countries.

2.3 Model

We introduce a model that can rationalize the empirical findings in Section 3.2. The model economy comprises a continuum of small open economies in a monetary union. Each small open economy is indexed by $k \in [0, 1]$.[5] In each small open economy, households consume two types of commodities: goods (G) and services (S). Goods can be exchanged frictionlessly among countries in the monetary union, while the services are produced and consumed exclusively within the domestic economy. Household preferences over the two commodities are non-homothetic: Goods are a necessity, and services are a luxury, consistent with the micro-level evidence presented in Section 3.2. Heterogeneous households inhabit each small open economy, differing in income, wealth, and their ability to share risks in financial markets. Countries differ in their distribution of income. Firms produce goods subject to nominal rigidities, implying that monetary policy has real effects.

2.3.1 Households

Two types of households inhabit each small open economy. Ricardian (R) households have access to financial markets and can borrow or save. Hand-to-Mouth (H) households cannot hold assets or debt. Each small open economy is inhabited by one representative household of each type. These two household types, each with a mass of one, are indexed by $j \in \{R, H\}$.

All households have identical preferences, independently of their type. Their preferences over a stream of expenditures $\{e_{j,t}\}_{t=0}^{\infty}$ are represented by:

$$V_{j,0} = \sum_{t=0}^{\infty} \beta^t v(e_{j,t}, P_t^G, P_t^S),$$

where $v(e_{j,t}, P_t^G, P_t^S)$ is an indirect utility function defined over total expenditure within the period $e_{j,t}$ and nominal prices of goods and services P_t^G, P_t^S. Prices are denominated

[5] In what follows, we suppress the country index k to lighten notation where describing properties of individual countries.

in units of currency, which serves as the numéraire of the economy. Expenditure satisfies $e_{j,t} = P_t^G c_{j,t}^G + P_t^S c_{j,t}^S$.

We adopt the functional form for the indirect utility function from Boppart (2014):

$$v(e_{j,t}, P_t^G, P_t^S) = \frac{1}{\varepsilon}\left[\left(\frac{e_{j,t}}{P_t^S}\right)^{\varepsilon} - 1\right] - \frac{\nu}{\gamma}\left[\left(\frac{P_t^G}{P_t^S}\right)^{\gamma} - 1\right]. \tag{2.4}$$

The parameters $\varepsilon, \gamma \in [0, 1]$ and $\nu > 0$. These preferences allow for non-homotheticity, and thus for expenditure shares over the two commodities that vary in the level of total expenditure. The parameter ε characterizes the degree of non-homotheticity. For $\varepsilon > 0$, the expenditure share on goods is decreasing in the level of total expenditure. The parameter ν controls the share of expenditure that is allocated to goods. γ governs the elasticity of substitution between the two commodities, which is not constant for $\gamma \neq \varepsilon$. These preferences embed homothetic, Cobb-Douglas preferences with $\varepsilon = \gamma = 0$.

Households supply labor to firms inelastically. Each household's labor endowment is denoted by l_j. Due to a nominal wage rigidity, some of this endowment may remain unemployed. $l_{j,t}$ denotes the amount of labor supplied by each household type that is employed by firms in a given period t.

Each household's budget constraint, expressed in units of currency, is:

$$P_t^G c_{j,t}^G + P_t^S c_{j,t}^S = W_{j,t} l_{j,t} + \Psi_{j,t} + P_t^G R_{t-1} b_{j,t} + R_{t-1}^m b_{j,t}^m - P_t^G b_{j,t+1} - b_{j,t+1}^m, \tag{2.5}$$

where the left-hand side represents expenditure $e_{j,t}$ and the right-hand side represents resources available to households. Labor income equals the product of the household-specific wage $W_{j,t}$ and the household's labor employment $l_{j,t}$. Nominal firms' profits accruing to household j are $\Psi_{j,t}$. Total income of each household thus equals $i_{j,t} = W_{j,t} l_{j,t} + \Psi_{j,t}$. Ricardian household resources additionally include holdings of real and nominal risk-free bonds $b_{R,t}$ and $b_{R,t}^m$, which yield gross returns R_{t-1} and R_{t-1}^m, respectively, net of purchases of real and nominal bonds $b_{j,t+1}$ and $b_{j,t+1}^m$. Hand-to-Mouth households do not hold bonds, $b_{H,t} = b_{H,t}^m = 0$.

Households allocate their expenditures across goods and services in each period as follows:[6]

$$c_{j,t}^G = \frac{e_{j,t}}{P_t^G}\left[\nu\left(\frac{P_t^S}{e_{j,t}}\right)^{\varepsilon}\left(\frac{P_t^G}{P_t^S}\right)^{\gamma}\right], \tag{2.6}$$

$$c_{j,t}^S = \frac{e_{j,t}}{P_t^S}\left[1 - \nu\left(\frac{P_t^S}{e_{j,t}}\right)^{\varepsilon}\left(\frac{P_t^G}{P_t^S}\right)^{\gamma}\right], \tag{2.7}$$

and the corresponding expenditure shares $\omega_{j,t}^G$ and $\omega_{j,t}^S$ are given by:

$$\omega_{j,t}^G \equiv \frac{P_t^G c_t^G}{e_{j,t}} = \nu\left(\frac{P_t^S}{e_{i,j,t}}\right)^{\varepsilon}\left(\frac{P_t^G}{P_t^S}\right)^{\gamma}, \tag{2.8}$$

$$\omega_{j,t}^S \equiv \frac{P_t^S c_t^S}{e_{j,t}} = 1 - \nu\left(\frac{P_t^S}{e_{j,t}}\right)^{\varepsilon}\left(\frac{P_t^G}{P_t^S}\right)^{\gamma}. \tag{2.9}$$

The share of expenditure allocated to goods decreases with total expenditure, since preferences in (2.4) imply that goods are a necessity. Conversely, the share allocated to services increases with total expenditure. This dependence of expenditure shares on total

[6]The household's intra-temporal problem associated with the indirect utility function (2.4) determines this allocation.

expenditure is a key feature of the model, marking a significant departure from most contributions to the heterogeneous-agents New Keynesian literature. With $\varepsilon = \gamma = 0$, the two expenditure shares are constant and given by ν and $1-\nu$ for goods and services, respectively, embedding the standard Cobb-Douglas case.

Ricardian households decide how to best allocate their expenditure over time. They maximize lifetime welfare subject to the stream of period budget constraints (2.5) and to standard transversality conditions. An inter-temporal Euler equation for expenditure describes how Ricardian households optimally allocate resources across periods:

$$\frac{v_e(e_{j,t}, P_t^G, P_t^S)}{v_e(e_{j,t+1}, P_{t+1}^G, P_{t+1}^S)} = \left(\frac{e_{j,t+1}}{e_{j,t}}\right)^{1-\varepsilon} \left(\frac{P_{t+1}^S}{P_t^S}\right)^{\varepsilon} = \beta R_t^m, \tag{2.10}$$

From this Euler equation and from the intra-temporal allocation of expenditure, the Ricardian household's intertemporal Euler equation for goods follows:

$$\frac{c_{R,t+1}^G}{c_{R,t}^G} = \beta R_t^m \left(\frac{P_t^G}{P_{t+1}^G}\right)^{1-\gamma} \left(\frac{P_t^S}{P_{t+1}^S}\right)^{\gamma}. \tag{2.11}$$

Arbitrage ensures equality of returns on nominal and real bonds, leading to the following Fisher equation:

$$R_t^m = R_t \frac{P_{t+1}^G}{P_t^G}. \tag{2.12}$$

Hand-to-Mouth households spend their entire labor and profit income in each period:

$$e_{H,t} = W_{H,t} l_{H,t} + \Psi_{H,t}.$$

2.3.2 Firms, Production, and Nominal Rigidities

Technology

A continuum of infinitesimal firms that produce either goods or services inhabits each small open economy.

Firms in each sector combine labor supplied by both Ricardian and Hand-to-Mouth households to produce output. The labor bundle, l_t^a, is the only factor used to produce each commodity $a \in \{G, S\}$, according to the production function:

$$y_t^a = (l_t^a)^{\alpha_a} \tag{2.13}$$

where α_a denotes the elasticity of output to labor in each sector s. The labor input bundle l_t^a is a Cobb-Douglas aggregator of labor supplied by Ricardian and Hand-to-Mouth households:

$$l_t^a \equiv \left(l_{H,t}^a\right)^{1-\xi_a} \left(l_{R,t}^a\right)^{\xi_a}, \tag{2.14}$$

where $l_{j,t}^a$ is the amount of labor supplied by household j that is employed in sector s. The parameter ξ_a controls the elasticity of the labor input bundle with respect to Ricardian households' labor in sector a. It also determines the share of total labor income in sector a accruing to Ricardian households.

Each household can supply labor to any firm, so labor of each type is mobile across sectors.

Firms' Problem and Distribution of Profits

Firms' maximize profits, defined as the revenue from sales minus the wage bill. The profit maximization problem for a representative firm in sector a is:

$$\max_{l_t^a} P_t^a \left(l_t^a\right)^{\alpha_a} - W_t^a l_t^a,$$

where P_t^a is the price of output, W_t^a is the wage associated with the labor bundle l_t^a.

The wage W_t^a is a Cobb-Douglas aggregator of the wages paid to Ricardian and Hand-to-Mouth households:

$$W_t^a = \left(\frac{W_{H,t}}{1-\xi_a}\right)^{1-\xi_a} \left(\frac{W_{R,t}}{\xi_a}\right)^{\xi_a},$$

where $W_{H,t}$ and $W_{R,t}$ denote the wages of Hand-to-Mouth and Ricardian households, respectively.

Profit maximization leads to the following labor demand schedules for each household type:

$$\alpha_a(1-\xi_a)P_t^a \frac{y_t^a}{l_{H,t}^a} = W_{H,t}, \tag{2.15}$$

and

$$\alpha_a \xi_a P_t^a \frac{y_t^a}{l_{R,t}^a} = W_{R,t}. \tag{2.16}$$

The total profits in sector a are given by:

$$\Psi_t^a = (1-\alpha_a)\, P_t^a y_t^a,$$

where profits reflect the share of output not allocated to labor costs.

Both Ricardian and Hand-to-Mouth households own equity in firms. Ownership is proportional to their contributions to labor income in each sector. Ricardian households own a fraction ξ_a of the equity in sector a, while Hand-to-Mouth households own $1-\xi_a$. Thus, profits are distributed as follows:

$$\Psi_{R,t} = \xi_G \Psi_t^G + \xi_S \Psi_t^S, \Psi_{H,t} = (1-\xi_G)\Psi_t^G + (1-\xi_S)\Psi_t^S.$$

This distribution implies that both household types share in the profits generated by firms in proportion to their labor contributions.

Nominal Rigidities

The wages of both worker types are constrained by a time-varying nominal rigidity:

$$W_{j,t} \geq \bar{\omega}_t W_{j,t-1},$$

where $\bar{\omega}_t$ controls the severity of downward wage rigidity.

When the wage rigidity binds, the labor demand schedules (2.15) and (2.16) determine employment for each worker type.

Total Real and Nominal Output

Total nominal output in each country equals the sum of the sectoral outputs, valued at their current-period nominal prices:

$$y_t^{nom} = P_t^G y_t^G + P_t^S y_t^S.$$

Total real output is similarly defined, with both goods and service outputs valued at base-year prices:

$$y_t = P_0^G y_t^G + P_0^S y_t^S.$$

2.3.3 Monetary Policy

A central bank controls the nominal interest rate R_t^m in the monetary union. It sets R_t^m to achieve a constant rate of inflation for the geometric average of the prices of the two commodities in the union:

$$\bar{\Pi} = \left(\frac{P_t^G}{P_{t-1}^G}\right)^{1-\gamma} \left(\frac{\tilde{P}_t^S}{\tilde{P}_{t-1}^S}\right)^{\gamma} = 1, \tag{2.17}$$

where the geometric weight γ coincides with the preference parameter. $\tilde{P}_t^S$ denotes the geometric average price of services across countries in the union:

$$\tilde{P}_t^S \equiv \exp\left(\int_0^1 \log P_{t,k}^S dk\right).$$

The central bank controls the supply of nominal bonds to maintain monetary policy objectives.

2.3.4 Market Clearing and Equilibrium

Labor market clearing ensures that the total labor employed by firms in both sectors does not exceed households' total labor endowment:

$$l_{j,t}^S + l_{j,t}^G = l_{j,t} \leq l_j \text{ for } j \in \{H, R\}, \tag{2.18}$$

where the inequality allows for involuntary unemployment arising from nominal wage rigidity. In each period, either all labor is fully employed, or the downward wage rigidity binds:

$$(l_{j,t}^S + l_{j,t}^G - l_{j,t})(W_{j,t} - \bar{\varpi}_t W_{j,t-1}) = 0 \text{ for } j \in \{H, R\}. \tag{2.19}$$

The market-clearing condition for services in each country k requires that consumption equals output:

$$c_{R,k,t}^S + c_{H,k,t}^S = y_{k,t}^S. \tag{2.20}$$

For each small open economy k in the monetary union, aggregate consumption of goods equals output, net of changes in net foreign assets:

$$c_{k,t}^G \equiv c_{R,k,t}^G + c_{H,k,t}^G = y_{k,t}^G + R_{t-1} b_{k,t} - b_{k,t+1}, \tag{2.21}$$

This expression can be rearranged to express the law of motion for the net foreign assets of country k, defining its current account:

$$ca_{k,t} \equiv b_{k,t+1} - b_{k,t} = y_{k,t}^G - c_{k,t}^G + b_{k,t}(R_{t-1} - 1) \tag{2.22}$$

The current account is given by the sum of the trade balance, $y^G_{k,t} - c^G_{k,t}$, and net interest payments on the stock of net foreign assets owned by the country at the start of the period, $b_{k,t}(R_{t-1} - 1)$.

If considering a small open economy in partial equilibrium, the path of $b_{k,t}$ can take any value, and thus $y^G_{k,t}$ and $c^G_{k,t}$ can diverge. If, instead, considering the interaction of all small open economies in the monetary union, market clearing for tradable goods at the union-level must hold:

$$\int_k y^G_{k,t} dk = \int_k (c^G_{R,k,t} + c^G_{H,k,t}) dk. \tag{2.23}$$

The no-arbitrage condition (2.12) implies that households are indifferent between holding real and nominal bonds. The net amount of nominal bonds supplied by the monetary authority is zero:

$$\int_k b^m_{R,k,t} dk = 0 \quad \text{for all } t. \tag{2.24}$$

Finally, in every period the world demand of the tradable good has to be equal the world production, $\int_0^1 c^G_{i,t} di = \int_0^1 y^G_{i,t} di$. This equilibrium condition implies that bonds are in zero net supply at the world level:

$$\int_0^1 B_{i,t+1} di = 0. \tag{2.25}$$

We are now ready to define a competitive equilibrium for a small open economy in isolation.

[Small Open Economy] Given a path for $\{R_t, P^G_t, R^m_t\}_t$, nominal rigidities $\{\bar{\varpi}_t\}_t$, and an initial condition b_0, a competitive equilibrium for country k consists of a path of quantities $\{c^G_{H,k,t}, c^G_{R,k,t}, c^S_{H,k,t}, c^S_{R,k,t}, e_{H,k,t}, e_{R,k,t}, l^G_{k,t}, l^G_{H,k,t}, l^G_{R,k,t}, l^S_{k,t}, l^S_{H,k,t}, l^S_{R,k,t}, y^G_{k,t}, y^S_{k,t}, b_{k,t+1}\}_t$, and prices $\{P^S_{k,t}, W_{H,k,t}, W_{R,k,t}\}_t$ satisfying households' optimality conditions (2.5), (2.6), (2.7), (2.10), the firm's profit maximization (2.13), (2.14), (2.15), (2.16), nominal rigidities (2.19), and the market-clearing condition (2.20).

An equilibrium for the monetary union is defined below.

[Monetary Union] Given an initial condition for $b^n_{R,k,0}$ for each country k and a path for nominal rigidities $\{\bar{\varpi}_t\}_t$, a competitive equilibrium consists of a path of quantities $\{c^G_{H,t}, c^G_{R,t}, c^S_{H,t}, c^S_{R,t}, e_{H,t}, e_{R,t}, l^G_t, l^G_{H,t}, l^G_{R,t}, l^S_t, l^S_{H,t}, l^S_{R,t}, y^G_t, y^S_t, b_{t+1}\}_t$, and prices $\{P^S_t, P^G_t, W_{H,t}, W_{R,t}, R_t, R^m_t\}_t$, satisfying for each country k the set of conditions in Equilibrium 2.3.4, the tradable goods market-clearing condition (2.23), the monetary policy target (2.17), and the Fisher equation (2.12).

Finally, we define a stationary allocation as an equilibrium of the model economy where all variables take constant values.[7]

2.4 Results: Revealing the Mechanism

This section leverages our model framework to explore how differences in expenditure composition across countries impact the strength of monetary policy transmission. Our main theoretical prediction, mirroring our empirical findings, is that in countries where the service expenditure share is higher, output responds less strongly to monetary policy

[7] Appendix 2.7 characterizes this stationary allocation in detail.

shocks. We build up to this result through a sequence of findings, first examining the determinants of expenditure shares and then how monetary policy transmission changes with varying degrees of inequality. Our headline findings rely on the following maintained assumptions:

Household preferences are non-homothetic, that is $\varepsilon > 0$, and that income effects dominate substitution effects $\gamma < \varepsilon$. The income in the service sector accrues disproportionately to Ricardian households, that is, $\xi_{S,k} > 0.5$ in all countries. The income in the goods sector accrues equally across households, $\xi_{G,k} = 0.5$, in all countries. We also explain how our qualitative results change when either of these two assumptions is relaxed. For quantitative illustrations of results from the model, we set the labor endowment of both households l_j to 0.93. This value normalizes expenditure to unity in a benchmark allocation where income shares are identical across households and sectors ($\xi_{G,k} = \xi_{S,k} = 0.5$) in all countries. We set the parameter governing the expenditure share ν to 0.6, implying an expenditure share on services in this benchmark allocation of 40%, which approximates the median of households' expenditure shares documented in Section 3.2. We set the parameter governing the degree of non-homotheticity ε to 0.18, following Boppart (2014). For analytical convenience, we set the remaining preference parameter $\gamma = 0$ in the baseline case but also consider a version of the model with $\gamma = \varepsilon$, which ensures a constant elasticity of substitution across goods below unity. We set the labor share of income in the service sector α_N to unity. In the goods sector, we set the labor share α_T to 0.43. These two parameters jointly determine an aggregate labor share of income of 0.66, consistent with standard stylized facts in macroeconomics.

2.4.1 Non-Homotheticity, Inequality, and Expenditure Shares

First, we examine the drivers of expenditure shares and show that households and countries with higher expenditure levels allocate a larger share of their expenditure to services. Our model's ability to generate heterogeneity in expenditure shares lies with non-homothetic preferences a la Boppart (2014). These preferences, where services are luxury goods and goods are necessities, have direct and intuitive consequences for consumption patterns, which we establish formally in Lemmas 3-5.[8]

First, as individual household income and expenditure rise, non-homothetic preferences lead them to allocate a successively larger share of their spending to services (Lemma 3). This micro-level behavior directly underpins our empirical finding that services expenditures increase with household income. Figure 2.6 illustrates this: under non-homotheticity, the expenditure share on services slopes upward with expenditure, in contrast to the flat line under homothetic preferences.

Lemma 3. *Under Assumption 2.4, each household allocates to services a share of its expenditure that is increasing in the level of expenditure itself:*

$$\frac{\partial \omega^S_{j,k}\left(e_{j,k}\right)}{\partial e_{j,k}} > 0.$$

Sketch of Proof: Consider the expenditure share on services $\omega^S_{j,k}$, as defined in (2.9). Given $\nu > 0$ and non-homotheticity ($\varepsilon > 0$), it is straightforward to show that the service expenditure share of household j in country k increases with its expenditure $e_{j,k}$. The partial derivative of the expenditure share with respect to the expenditure level is

[8] Appendix 2.7 provides more details on the proofs for these Lemmas.

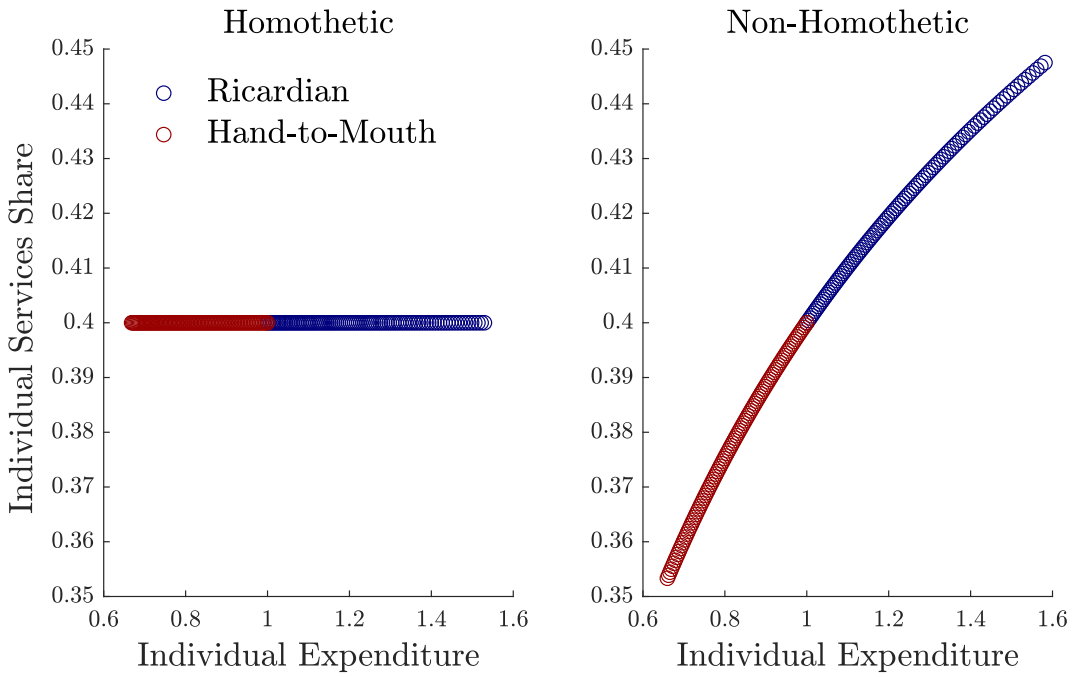


Figure 2.6: Household-level service expenditure share $\omega_{j,k}^S$ as a function of household expenditure.

Notes: Stationary allocation under homothetic ($\varepsilon = 0$) or non-homothetic ($\varepsilon > 0$) preferences. Each dot represents one household, either Ricardian or Hand-to-Mouth, in one of the countries of the monetary union. Countries in the monetary union differ in terms of their income distribution, with $\xi_{S,k} \in [0.5, 1]$ and $\xi_{T,k} = 0.5$.

given by:

$$\frac{\partial \omega_{j,k}^S\left(e_{j,k}\right)}{\partial e_{j,k}} = \varepsilon \frac{1-\omega_{j,k}^S}{e_{j,k}},$$

which is strictly positive for $\varepsilon > 0$. □

Aggregating this micro household behavior, two further implications emerge. First, countries with higher average expenditure levels will, on average, also have a larger share of services consumption:

Lemma 4. *Under Assumption 2.4, the average expenditure share allocated to services in a country, $\bar{\omega}_k^S$, increases with the country's aggregate expenditure level, $\bar{e}_k$:*

$$\frac{\partial \bar{\omega}_k^S\left(\bar{e}_k\right)}{\partial \bar{e}_k} > 0,$$

given prices and the distribution of households' relative expenditures, $\{x_{j,k}\}_{j\in\{H,R\}}$.

Sketch of Proof: This result follows directly from the behavior of household-level expenditure share established in Lemma 3. The average expenditure share satisfies

$$\bar{\omega}_k^S\left(\bar{e}_k\right) = x_{H,k}\omega_{H,k}^S\left(x_{H,k}\bar{e}_k\right) + x_{R,k}\omega_{R,k}^S\left(x_{R,k}\bar{e}_k\right) \tag{2.26}$$

As Lemma 3 establishes that household-level service expenditure shares increase with expenditure, the average service expenditure share also increases in aggregate expenditure, given the distribution of such expenditure across households. □

Second, and central for our analysis, countries with greater income inequality will also feature a higher average service expenditure share (Lemma 5). This occurs because higher inequality, driven by a larger share of service sector income accruing to wealthier

Ricardian households (Assumption 2.4), concentrates a larger portion of total expenditure among these Ricardian households. Given their higher income and non-homothetic preferences, these households spend disproportionately more on services, increasing the aggregate service share.[9] This mechanism provides a theoretical basis for our *Empirical Regularity 1* (service consumption increases with income inequality). Figure 2.7 demonstrates this positive relationship between the Gini index and the average service expenditure share.

Lemma 5. *Under Assumptions 2.4 and 2.4, the average expenditure share allocated to services in one country, $\bar{\omega}_k^S$, increases with the dispersion of household expenditure in that country, $\sigma_{x,k}$:*

$$\frac{\partial \bar{\omega}_k^S\left(\bar{e}_k\right)}{\partial \sigma_{x,k}} > 0,$$

Sketch of Proof: To see this result, consider the expression for the average expenditure share in (2.26). For each household, total service expenditure is increasing and convex in own expenditure $e_{j,k}$:

$$P^S c_{j,k}^S = e_{j,k} - \nu \left(e_{j,k}\right)^{1-\varepsilon} P^{S\varepsilon} \left(\frac{P_t^G}{P_t^S}\right)^{\gamma}.$$

The average service expenditure share is therefore a convex function of the households' shares of total expenditure $x_{j,k}$, resembling the household-level expenditure share. By Jensen's inequality, an increase in the dispersion of $(x_{H,k}, x_{R,k})$ away from equality $(0.5, 0.5)$ leads to a higher average service expenditure share $\bar{\omega}_k^S$. □

2.4.2 Monetary Policy Transmission: The Role of Income Inequality and Expenditure Shares

Having established how non-homothetic preferences and the income distribution shape expenditure patterns, we now turn to our central question: how do these features affect the transmission of monetary policy? We analyze a temporary, unexpected contractionary monetary policy shock (an increase in R_t^m at $t = 1$), assuming nominal wage rigidities bind only in the short run ($t = 1$).

Benchmark: A Monetary Union of Homogeneous Countries

To isolate the core mechanisms through which the income distribution and non-homothetic preferences influence monetary policy's impact, we first analyze a monetary union of identical countries—a benchmark equivalent to a closed economy. Here, the interaction between non-homothetic preferences and income inequality is key for understanding the dampened transmission of monetary policy.

Our main theoretical prediction for this homogeneous union (formalized in Proposition 6) is the comparative static that the higher the income inequality with the member countries of the monetary union, the weaker is the transmission of monetary policy to total output and, as established in Section 2.4.1 (Lemma 5), the higher is average expenditure share on services. This occurs when higher inequality stems from Ricardian households deriving a larger fraction ξ_S of their income from the service (luxury) sector

[9] Considering the opposite case with $\xi_{S,k} \in [0, 0.5]$ would give rise to higher inequality, the lower is $\xi_{S,k}$, but with the unusual assumption that the Hand-to-Mouth household is the one with the highest income.

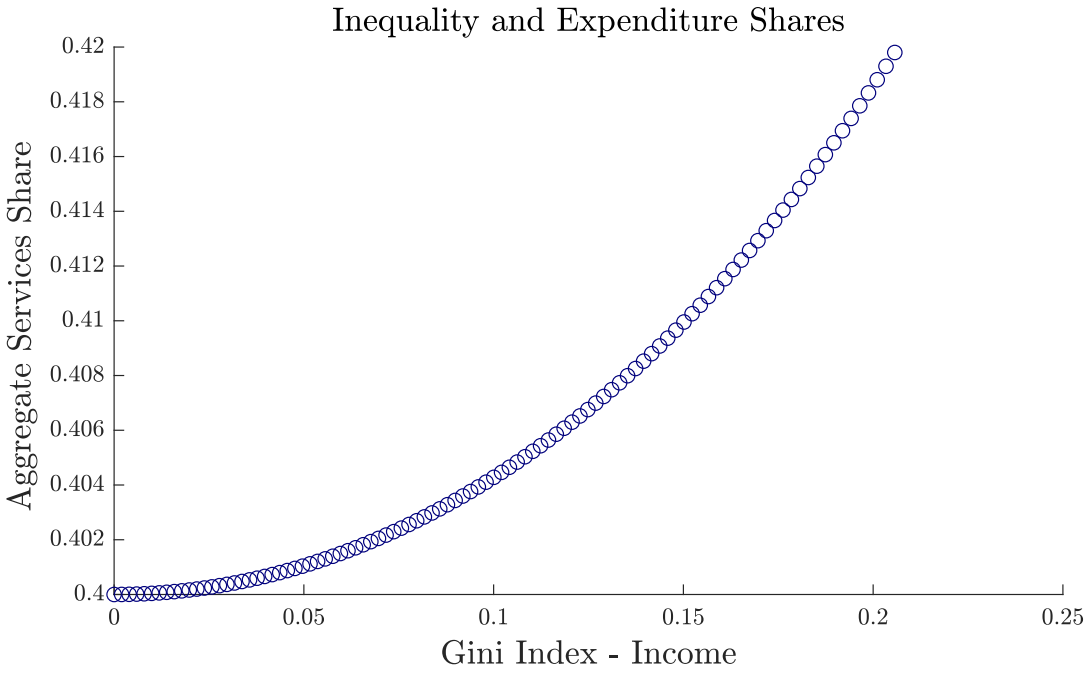


Figure 2.7: Average service expenditure share $\bar{\omega}_k^S$ as a function of income inequality.

Notes: Income inequality is represented by the Gini index on household income. Stationary allocation with non-homothetic preferences ($\varepsilon > 0$). Each dot represents one country in the monetary union. Countries differ in income distribution, with $\xi_{S,k} \in [0.5, 1]$ and $\xi_{T,k} = 0.5$.

(Assumption 2.4). This theoretical result directly rationalizes our primary *Empirical Regularity 2*: countries with high service expenditure shares experience a dampened output response to monetary policy. Figure 2.8 graphically depicts how the output response to a contractionary shock becomes more muted as the service expenditure share (driven by inequality ξ_S) rises.

Proposition 6. *Under Assumptions 2.4 and 2.4, countries feature: i) a weak transmission of monetary policy to output,*

$$\left.\frac{y_1}{y_2}\right|_{Unequal} > \left.\frac{y_1}{y_2}\right|_{Equal}$$

ii) a high share of average expenditure allocated to the service good,

$$\bar{\omega}^S_{Unequal} > \bar{\omega}^S_{Equal}.$$

Proof: See Appendix 2.7.

The intuition behind this muted transmission hinges on how income inequality itself responds to the policy shock, becoming *pro-cyclical*. Because services are luxuries, their relative value is pro-cyclical (falls in recessions, rises in expansions).[10] When Ricardian households' income is disproportionately tied to the service sector ($\xi_S > \xi_G$), their income also becomes more pro-cyclical than that of Hand-to-Mouth households.

Following a contractionary monetary policy shock:

1. The relative value of services falls. Ricardian households, disproportionally exposed, see a sharp income decline.

[10] Appendix 2.7 establishes that the relative value of service output is increasing in aggregate expenditure and therefore pro-cyclical under non-homothetic preferences.

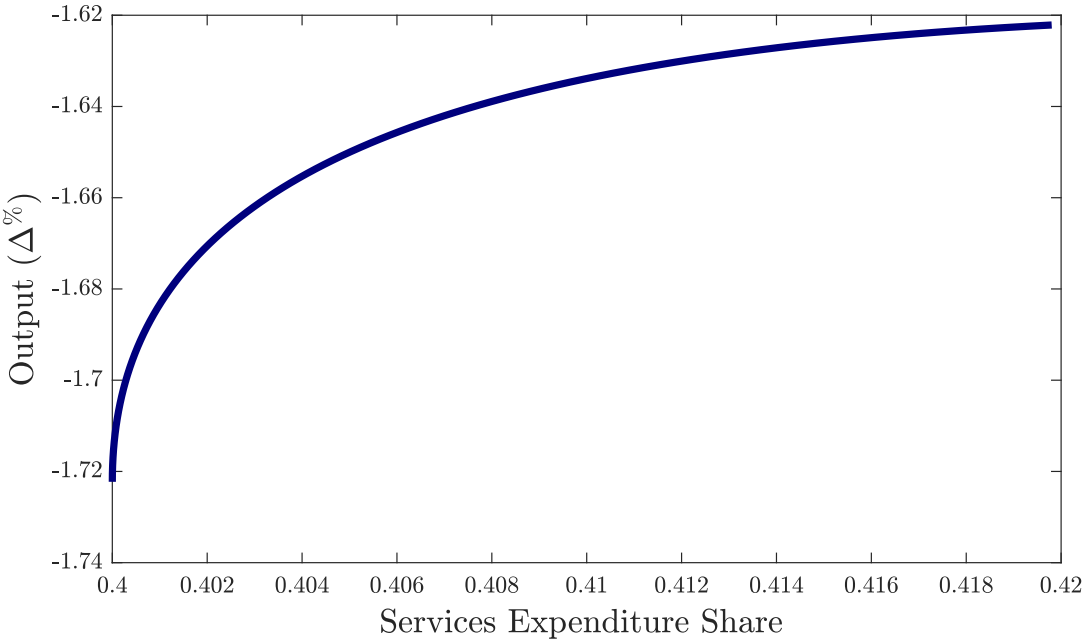


Figure 2.8: Contractionary monetary policy in union of identical countries. Effect on total output, as function of the stationary-allocation service share of expenditure.

Notes: The monetary policy shock amounts to a one-period increase in the nominal interest rate from 1% to 2%.

2. Conversely, Hand-to-Mouth households, less exposed to the service sector contraction, experience a *milder* income drop. This is formally established in Lemma 8, which shows that the Ricardian household's expenditure relative to the Hand-to-Mouth's (e_R/e_H) falls during the policy-induced recession (Figure 2.9, right panel).
3. This differential income response means that overall income inequality (between Ricardian and Hand-to-Mouth) actually *decreases* during the recession (hence that inequality is pro-cyclical).
4. Since Hand-to-Mouth consumption is highly sensitive to current income, their smaller income drop translates to a smaller consumption drop. This effect is crucial because, as shown in Lemma 9, the pro-cyclicality of e_R/e_H also makes the Ricardian share of total tradable consumption (c_R^G/c^G) pro-cyclical. Thus, during the contraction, Hand-to-Mouth households account for a larger share of tradable consumption.

This pro-cyclical behavior of inequality, driven by non-homothetic preferences and the specific income structure, effectively insulates Hand-to-Mouth consumption partially from the shock. Since these households have a high marginal propensity to consume, the overall drop in aggregate demand—particularly for tradable goods—is less acute.

Proposition 7 formalizes this by showing that the aggregate contraction in goods consumption (and thus output in this closed setting) is *weaker* than the direct fall in Ricardian households' tradable consumption demand. Ricardian demand falls proportionally to the shock itself (i.e., $\frac{c_{R,1}^G}{c_{R,2}^G} = \frac{\bar{\Pi}}{\beta R_1^m}$),[11] but the overall impact is dampened by the relatively stable consumption of Hand-to-Mouth agents. The left panel of Figure 2.9 illustrates this: the aggregate goods consumption contraction is milder than what Ricardian

[11] This follows from the Ricardian household's Euler equation for goods (2.11), total expenditure (2.10), intra-temporal allocation (2.8), and the inflation target (2.17). In a homogeneous union, $\tilde{P}_t^S = P_{k,t}^S$.

demand alone would imply.

Proposition 7. *Under Assumptions 2.4 and 2.4, a contractionary monetary policy shock reduces goods output and aggregate goods consumption less than it reduces the demand for goods by Ricardian households, which are directly exposed to the shock:*

$$\frac{c^G_{R,1}}{c^G_{R,2}} < \frac{y^G_1 = c^G_1}{y^G_2 = c^G_2}.$$

Proof: See Appendix 2.7.

Lemma 8. *The Ricardian household's relative expenditure is pro-cyclical, i.e.:*

$$\frac{e_{R,1}}{e_{H,1}} < \frac{e_{R,2}}{e_{H,2}},$$

if i) $\xi_S > \xi_G$*, and ii)* $\varepsilon > 0$.

Proof: See Appendix 2.7.

Lemma 9. *The Ricardian's share of total goods consumption* $c^G_{R,t}/c^G_t$ *is pro-cyclical, if and only if the expenditure of this household, relative to that of the Hand-to-Mouth,* $e_{R,t}/e_{H,t}$*, is pro-cyclical:*

$$\frac{e_{R,2}}{e_{H,2}} > \frac{e_{R,1}}{e_{H,1}} \iff \frac{c^G_{R,2}}{c^G_2} > \frac{c^G_{R,1}}{c^G_1}.$$

Proof: See Appendix 2.7.

Since service and total output move in lockstep with output of goods in this setting,[12] the weaker response of goods' output directly translates to a weaker response of total output (as shown in Proposition 6). Figure 2.10 further illustrates that the impact on goods' prices and service output is also weaker in more unequal countries. Both non-homothetic preferences and the specific income distribution ($\xi_S > \xi_G$) are crucial; without either, this pro-cyclical inequality channel would not operate, and goods' output would fall in line with Ricardian demand, as in a representative-agent setting.

We now transition to a setting with heterogeneous countries, where these core mechanisms interact with cross-country trade, allowing us to explore how heterogeneity in monetary policy transmission arises across countries within the union — our main empirical focus.

Extended Framework: Monetary Union of Heterogeneous Countries

We now extend our analysis to a monetary union where countries differ in their income distribution, specifically in the share of service income accruing to Ricardian households ($\xi_{S,Un} > \xi_{S,Eq}$, while ξ_G is uniform). This heterogeneity delivers two key results, formalized in Propositions 10 and 11.

First, consistent with the homogeneous case and our empirical findings, monetary policy exerts a weaker effect on total output in the "unequal" countries (those with higher ξ_S

[12] Appendix 2.7 proves this result. This co-movement arises because output of goods responds positively to its price, and service output (a luxury) increases with both goods' output and the relative price of goods.

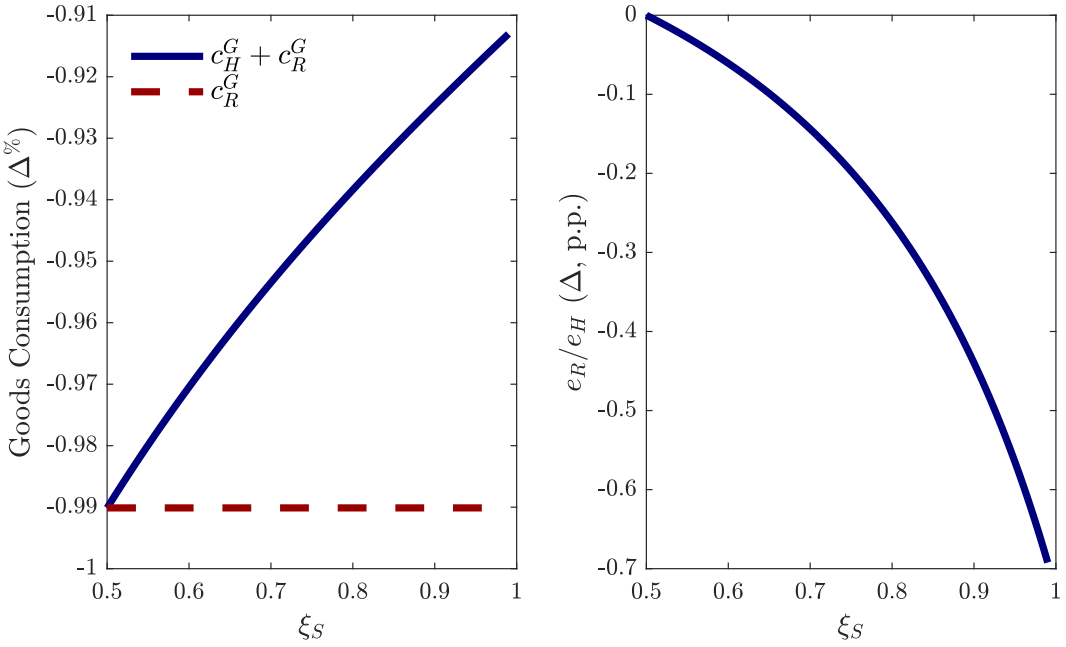


Figure 2.9: Contractionary monetary policy in union of identical countries. Effects of monetary policy shock on consumption of goods, aggregate and of Ricardian household (left panel) and on relative expenditure of Ricardian household (right panel).

Notes: Effects are displayed as function of the share of service sector output accruing to Ricardian household (ξ_S). The monetary policy shock amounts to a one-period increase in the nominal interest rate from 1% to 2%.

and thus higher service expenditure shares) compared to the "equal" countries (Proposition 10). Figure 2.11 (left panel) illustrates how the output decline in unequal countries is less pronounced.

Proposition 10. *In a monetary union of heterogeneous countries, monetary policy has a weaker effect on total output in "unequal" than in "equal" countries. Upon a contractionary shock:*

$$\frac{y_{1,Un}}{y_{0,Un}} > \frac{y_{1,Eq}}{y_{0,Eq}}.$$

Second, and novel to the heterogeneous country setting, monetary policy shocks generate systematic trade imbalances: unequal countries run trade deficits, while equal countries run trade surpluses following a contractionary shock (Proposition 11). This rationalizes our empirical observations (Empirical Regularity 3) that trade flows respond heterogeneously to monetary policy. Figure 2.11 (right panel) shows the emergence of these imbalances.

Proposition 11. *In a monetary union of heterogeneous countries, a contractionary monetary policy shock gives rise to a trade deficit in the "unequal" countries and to a trade surplus in the "equal" ones:*

$$\frac{c^G_{1,Un}}{y^G_{1,Un}} > \frac{c^G_{2,Un}}{y^G_{2,Un}} \quad \textit{and} \quad \frac{c^G_{1,Eq}}{y^G_{1,Eq}} < \frac{c^G_{2,Eq}}{y^G_{2,Eq}}.$$

The mechanics behind these cross-country differences (detailed below Lemma 12 revolve around the common price of tradables ($P^G_{1,MU}$) in the union. This union-wide price settles between what would be the autarky prices for equal and unequal countries (Lemma 12).

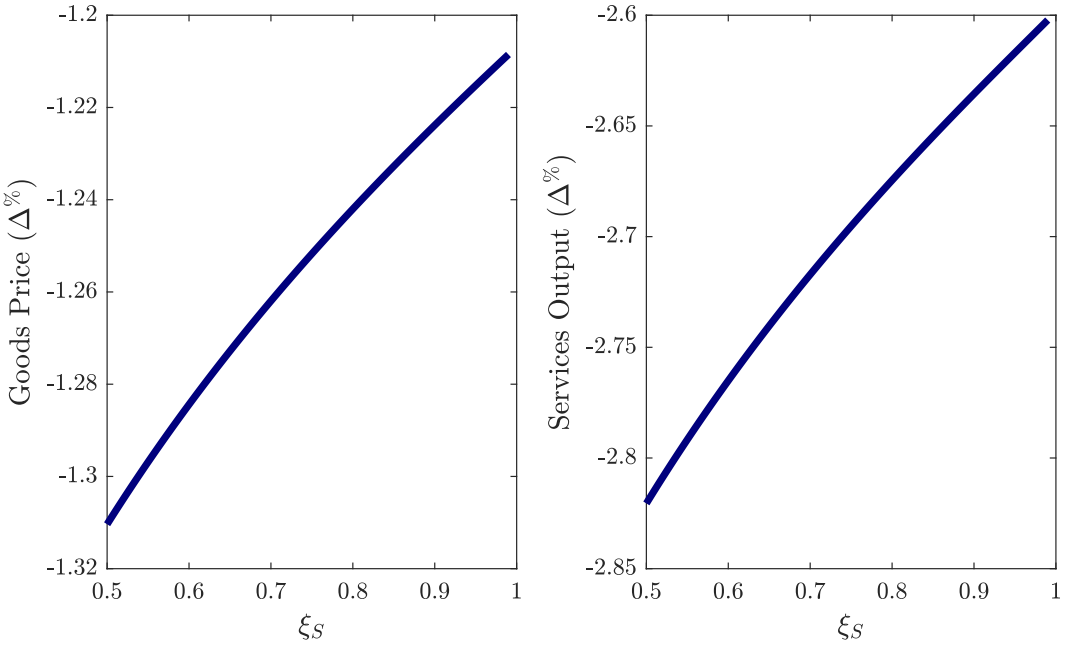


Figure 2.10: Contractionary monetary policy in union of identical countries. Effects on the price of tradable good (left panel) and on service output (right panel).

Notes: Effects are displayed as function of the share of service sector output accruing to Ricardian household (ξ_S). The monetary policy shock amounts to a one-period increase in the nominal interest rate from 1% to 2%.

While goods' output declines similarly across all countries due to the common shock and price, total goods' *consumption* declines *less* in unequal countries. This is because the pro-cyclical inequality mechanism (milder income fall for Hand-to-Mouth agents) still operates, supporting their consumption more. To finance this relatively higher consumption, unequal countries must import more goods (or export less), leading to a trade deficit. Conversely, equal countries, lacking this dampening mechanism, see a larger fall in consumption and thus run trade surpluses. This differential impact on goods consumption also means the contraction in service output (and thus total output) is smaller in unequal countries, reinforcing Proposition 10.

Lemma 12. *In a monetary union of heterogeneous countries, the equilibrium price of goods in the initial period is higher than the closed-economy equilibrium price of the equal countries and is lower than that of the unequal countries:*

$$P^G_{1,Eq,CE} < P^G_{1,MU} < P^G_{1,Un,CE}.$$

2.4.3 Model Robustness and Empirical Grounding

To bolster confidence in our findings, we conduct three sets of exercises. First, we test the sensitivity of our theoretical results to key modeling assumptions. Second, we empirically investigate a key underpinning of our model: the link between aggregate inequality and the sectoral income distribution of Ricardian households. Third, we present extensive robustness checks for our primary empirical facts in Appendix A.7.

Sensitivity of Model Results to Assumptions

Our core theoretical predictions prove robust to a range of alternative modeling choices.

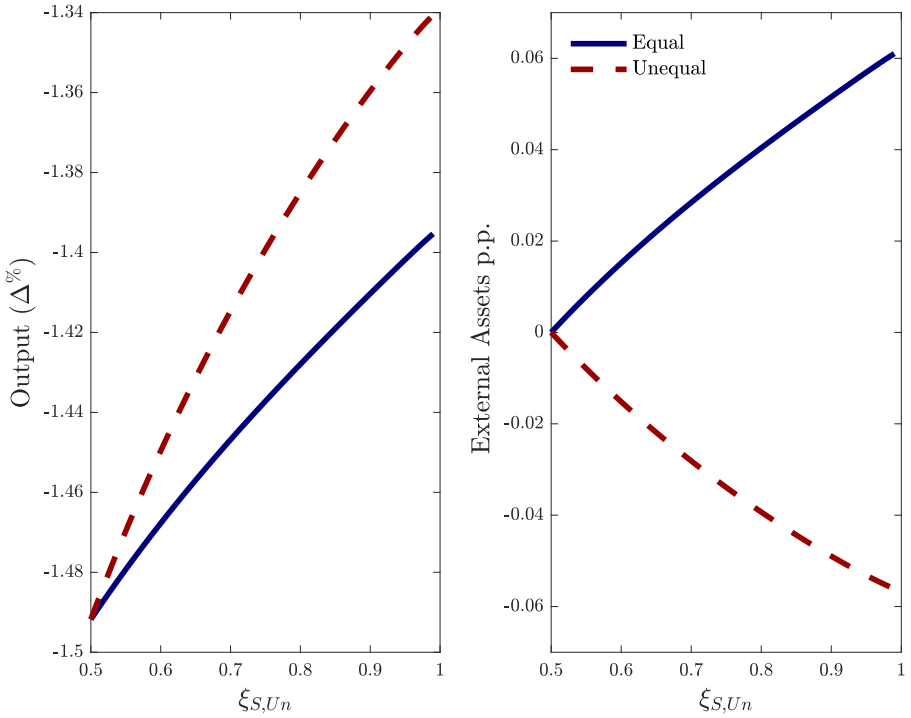


Figure 2.11: Contractionary monetary policy in union of heterogeneous countries. Effects of monetary policy shock on total output and trade balance.

Notes: Effects for Equal and Unequal countries are displayed as function of the share of service sector output accruing to the Ricardian household in the Unequal countries ($\xi_{S,Un}$). The monetary policy shock amounts to a one-period increase in the nominal interest rate from 1% to 2%.

Elasticity of Substitution between goods and services: Varying the elasticity of substitution (Figure 2.12) reveals that while the *magnitude* of monetary policy's impact changes, the *qualitative* result—that higher service income for Ricardians ($\xi_{S,Un}$) dampens policy transmission—holds unless the goods are extremely poor substitutes. Milder fluctuations in the relative demand for services (at lower elasticities) reduce the pro-cyclicality of inequality, thereby diminishing, but not eliminating, the dampening effect.

Degree of Non-Homotheticity: A higher degree of non-homotheticity (a larger ϵ, Figure 2.13) amplifies the overall impact of monetary policy, as demand for luxury services becomes even more sensitive to income changes. Simultaneously, it *strengthens* our main finding: the difference in output response between unequal and equal countries becomes more pronounced because the pro-cyclicality of inequality is magnified.

Source of Income Inequality: The specific *source* of income inequality is paramount. If inequality stemmed primarily from Ricardian households earning more from the *goods* sector (differing ξ_G rather than ξ_S), monetary policy effects would actually be *stronger* in unequal countries (Figure 2.4.3). This underscores the importance of our model's assumption, which is supported by our empirical analysis in Section 2.4.3 that inequality is largely driven by Ricardian exposure to the service sector. Results remain consistent even if goods sector income is also unequal, as long as the *differential* cyclicality is driven by service income (Figure 2.15).

Richer Household Heterogeneity: While our two-agent framework captures the essential dynamics, extending it to a continuum of agents where sectoral income exposures emerge endogenously from differences in wealth or risk aversion is a promising avenue. Such a model could endogenize the income distributions (ξ_S, ξ_G) we exogenously vary and

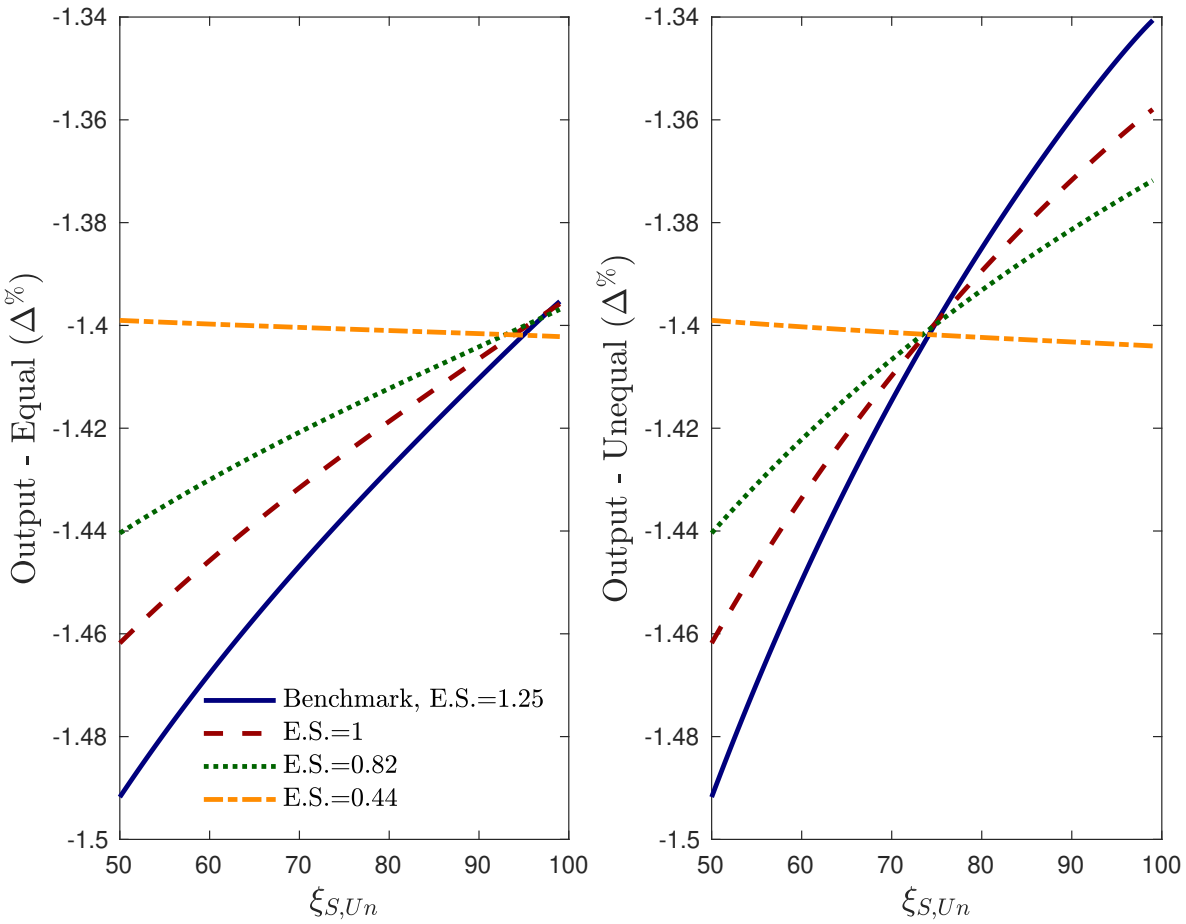


Figure 2.12: Contractionary monetary policy in union of heterogeneous countries. Alternative values for elasticity of substitution.

Notes: Effects of monetary policy shock on total output in Equal and Unequal countries, displayed as function of the share of service sector output accruing to the Ricardian household in the Unequal countries ($\xi_{S,Un}$). Each of the different lines correspond to a different value for the elasticity of substitution between tradable and services. The monetary policy shock amounts to a one-period increase in the nominal interest rate from 1% to 2%.

introduce other heterogeneous responses, like precautionary savings, further enriching the analysis of monetary policy transmission.

Empirical Validation of Sectoral Income Distribution Assumption

A cornerstone of our model's ability to explain heterogeneous monetary policy effects is the assumption that cross-country differences in income inequality are linked to Ricardian households earning a larger share of their income from the service sector, especially in more unequal countries. We empirically test this in Section 3.2. Our findings there confirm that higher income inequality (e.g., higher Gini) is indeed associated with a larger difference in the service income share between Ricardian and Hand-to-Mouth households, lending empirical support to this key model mechanism. (Further details on this empirical exercise are provided in Appendix A.3).

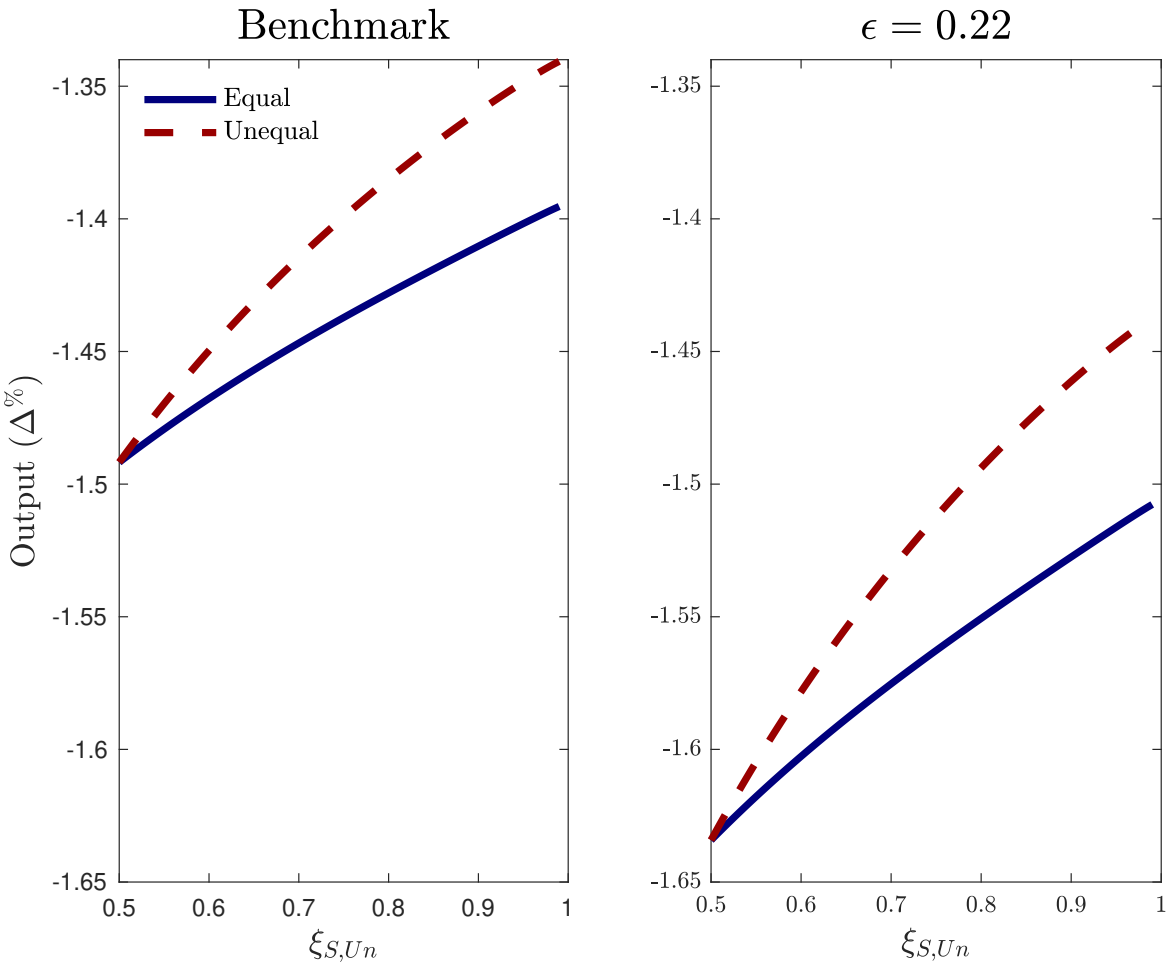


Figure 2.13: Contractionary monetary policy in union of heterogeneous countries. Higher degree of non-homotheticity.

Notes: Effects of monetary policy shock on total output in Equal and Unequal countries, displayed as function of the share of service sector output accruing to the Ricardian household in the Unequal countries ($\xi_{S,Un}$). Model with higher degree of non-homotheticity in preferences, $\epsilon = 0.22$, compared to the benchmark case with $\epsilon = 0.18$. The monetary policy shock amounts to a one-period increase in the nominal interest rate from 1% to 2%.

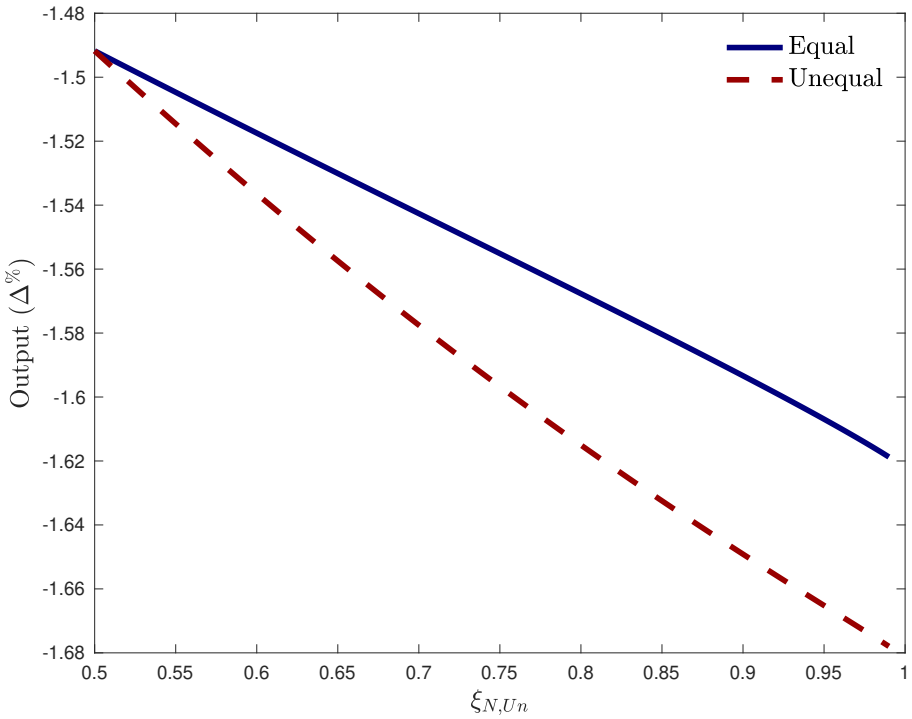


Figure 2.14: Contractionary monetary policy in union of heterogeneous countries. Inequality stemming from goods sector.

Effects of monetary policy shock on total output in Equal and Unequal countries, displayed as function of the share of tradable sector output accruing to the Ricardian household in the Unequal countries ($\xi_{G,Un}$). In this setting with alternative income distribution, income from the service sector is distributed equally in all countries: $\xi_{S,Un} = \xi_{S,Eq} = 0.5$. The monetary policy shock amounts to a one-period increase in the nominal interest rate from 1% to 2%.

Notes: Effects of monetary policy shock on total output in Equal and Unequal countries, displayed as function of the share of tradable sector output accruing to the Ricardian household in the Unequal countries ($\xi_{G,Un}$). In this setting with alternative income distribution, income from the service sector is distributed equally in all countries: $\xi_{S,Un} = \xi_{S,Eq} = 0.5$. The monetary policy shock amounts to a one-period increase in the nominal interest rate from 1% to 2%.

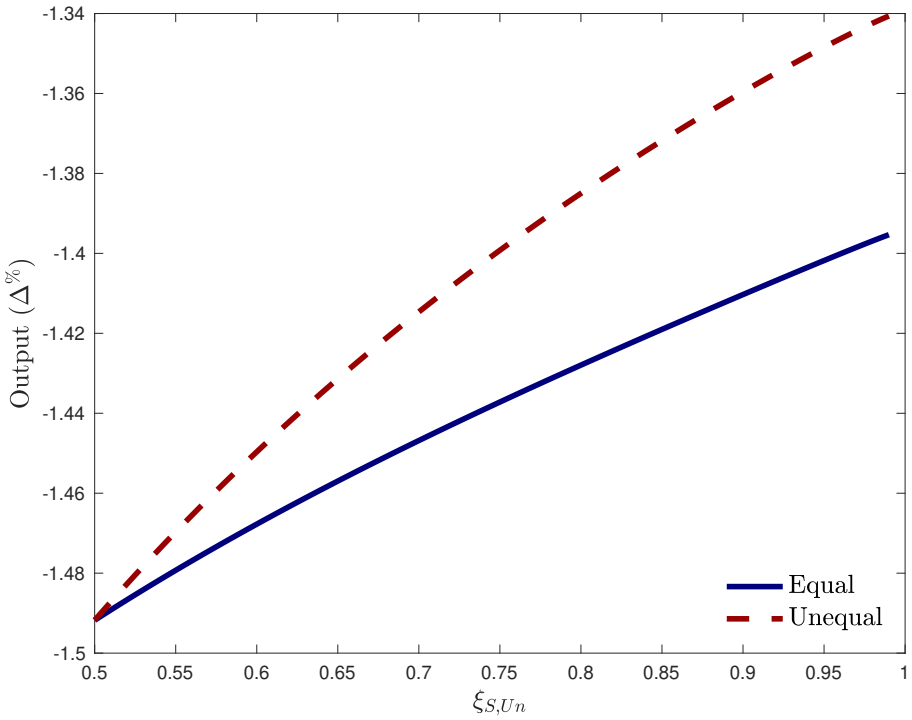


Figure 2.15: Contractionary monetary policy in union of heterogeneous countries. Inequality in both sectors.

Notes: Effects of monetary policy shock on total output in Equal and Unequal countries, displayed as function of the share of service sector output accruing to the Ricardian household in the Unequal countries ($\xi_{S,Un}$). Under this distribution, income from the goods sector is distributed unequally in all countries: $\xi_{G,Un} = \xi_{G,Eq} = 0.75$. The monetary policy shock amounts to a one-period increase in the nominal interest rate from 1% to 2%.

2.5 Conclusion

This paper examined how cross-country heterogeneity in income and its distribution influences the transmission of common monetary policy within a currency union, with a focus on the expenditure composition channel. We empirically established four key stylized facts for the euro area: (1) households with higher income dedicate a larger share of their expenditure to services (services); (2) countries with greater income inequality exhibit higher aggregate service expenditure shares; (3) countries with larger service expenditure shares show weaker output responses to monetary policy shocks (when wealthier households' disproportionate income from the service sector); and (4) these countries also tend to experience trade balance deteriorations following contractionary monetary policy, contrasting with improvements in low service share countries.

We rationalized these findings within a heterogeneous-agent New Keynesian model incorporating non-homothetic preferences and unequal sectoral income exposures. The model generates a weaker impact of monetary policy in high service share countries, despite greater price rigidities in that sector. The critical mechanism is *pro-cyclical inequality*: when wealthier (Ricardian) households are more exposed to the volatile luxury service sector, a contractionary shock reduces their income more than that of Hand-to-Mouth households. This relative income shift cushions the consumption of Hand-to-Mouth agents, dampening the overall aggregate demand contraction and thus the policy's impact on output. In a monetary union, this also leads to the observed heterogeneous trade balance responses.

Our results highlight that understanding the interplay between household income distribution, non-homothetic consumption patterns, and sectoral exposures is crucial for assessing the effectiveness of monetary policy in diverse economic unions. Future research could extend this framework to explore optimal monetary policy design in such settings and to endogenize households' sectoral income exposures for a richer treatment of inequality.

2.6 Appendix - Empirical Analysis

Services consumption shares by income and country

Figure 2.3 shows the average 2000-2020 service expenditure shares by disposable income across euro area member states. The data is based on the household budget survey (HBS). The figure shows that both across households as well as countries the share of services increases with income revealing non-homothetic preferences. Table 2.2 further summarizes the graph showing the slope of the linear fit as well as the average disposable income and service expenditure share for each country.

Country	Slope	Disposable Income	Services Expenditure share
Austria	0.09	22062	37.70
Belgium	0.17	20243	42.34
Cyprus	0.60	15904	43.70
Estonia	1.62	6367	27.64
Finland	0.13	22134	42.07
France	0.04	20904	42.69
Germany	0.25	20222	41.20
Greece	0.75	10674	42.58
Ireland	0.36	22281	44.34
Italy	0.30	15601	35.92
Latvia	2.20	4719	29.60
Lithuania	2.44	4363	24.49
Luxembourg	-0.08	34314	40.47
Malta	0.59	11775	36.33
Netherlands	0.01	20982	45.85
Portugal	0.86	9347	44.35
Slovakia	1.97	5570	28.69
Slovenia	0.42	11468	37.30
Spain	0.42	14175	42.18

Table 2.2: Summary statistics on countries income and consumption.
Notes: The slope corresponds to the linear fit of the service expenditure share on the level of disposable income at every income quintile.

Inequality and Consumption Baskets

We show that service expenditure share increase with income inequality at the country level. Table 2.3 shows the regression results of the following specification:

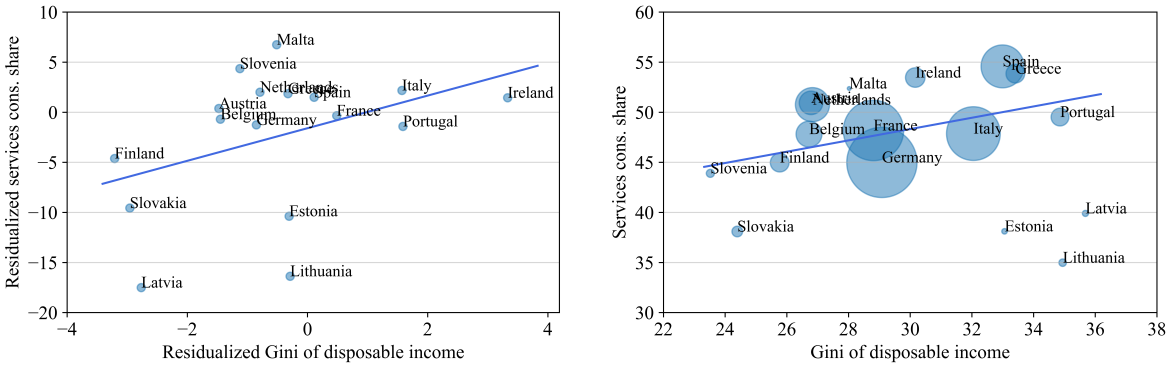
$$\omega_k = \alpha + \beta Gini_c + \gamma X'_k + + \epsilon_c$$

where ω_k is the average 2000-2020 service expenditure share of country k, and X' is a vector of controls averaged between 2000-2020 including the trade balance, the old-age dependency ratio, government expenditure and GDP per capita. Countries are weighted by there GDP respectively. Figure 2.16 further plots the results for the unweighted regression as well as the regression excluding controls.

	Services expenditure share		
Gini	1.046*		
	(-0.397)		
p90p10		6.128**	
		(-1.596)	
p75p25			13.36**
			(-3.91)
Trade balance	-1.034	-0.658	-0.769
	(-0.664)	(-0.529)	(-0.534)
Old-age dependency	-0.764**	-0.769***	-0.679**
	(-0.209)	(-0.173)	(-0.166)
Gov. exp.	-0.246	0.0512	-0.0482
	(-0.364)	(-0.288)	(-0.286)
GDP p.c.	0.648	0.625*	0.597*
	(-0.327)	(-0.231)	(-0.244)
N	17	17	17
adj. R-sq	0.571	0.689	0.644

Table 2.3: Regression table of equation 2.1.

Notes: All variables are averaged over 2000 to 2020. The variables p90p10 and p75p25 correspond to the income ratios of the 90th and 10th as well as the 75th and 25th percentile, respectively.



(a) Unweighted linear fit.

(b) Un-residualzed linear fit.

Figure 2.16: Inequality and service expenditure shares across countries.

Relationship between sectoral income shares and inequality

While countries' consumption baskets are linked to their income inequality, we further show the sectoral origins of income inequality. We first show how the share of income generated in service sectors increases with the income level. We then further discriminate between Ricardian and Hand-to-Mouth households to pin down the difference in sectoral income shares. We use the first three waves of HFCS data on household incomes and the sector of employment. Finally, we follow Almgren et al. (2022) in defining households as Ricardian or Hand-to-Mouth based on their wealth-income composition.

Figure 2.17 shows that the service share of income increases by income decile within and across EU members.

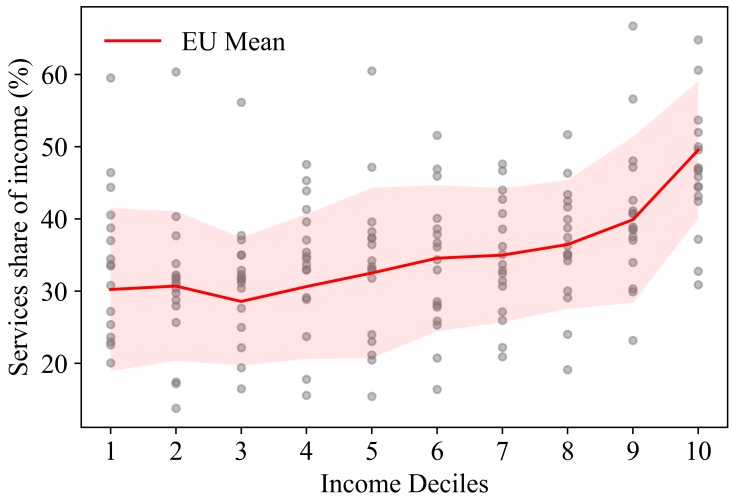


Figure 2.17: Service income shares by decile.

Notes: This graphs shows the service share of income by income decile across EU members and on average. The red shaded area represents one standard deviation.

COICOP consumption classification

Table 2.4 lists all three-digit COICOP consumption categories in the household budget survey and their classification as goods or services.

COICOP		Name	Goods
CP01		*Food and non-alcoholic beverages*	
	CP011	Food	Yes
	CP012	Non-alcoholic beverages	
CP02		*Alcoholic beverages, tobacco and narcotics*	Yes
	CP021	Alcoholic beverages	Yes
	CP022	Tobacco	Yes
CP03		*Clothing and footwear*	
	CP031	Clothing	Yes
	CP032	Footwear	Yes
CP04		*Housing, water, electricity, gas and other fuels*	
	CP041	Actual rents for housing	No
	CP042	Imputed rents for housing	No
	CP043	Maintenance and repair of the dwelling	No
	CP044	Water supply and miscellaneous services relating to the dwelling	No
	CP045	Electricity, gas and other fuels	Yes
CP05		*Furnishings, household equipment and routing household maintenance*	
	CP051	Furniture and furnishings, carpets and other floor coverings	Yes
	CP052	Household textiles	Yes
	CP053	Household appliances	Yes
	CP054	Glassware, tableware and household utensils	Yes
	CP055	Tools and equipment for house and garden	Yes
	CP056	Goods and services for routine household maintenance	Yes
CP06		*Health*	
	CP061	Medical products, appliances and equipment	Yes
	CP062	Out-patient services	No
	CP063	Hospital services	No
CP07		*Transport*	
	CP071	Purchase of vehicles	Yes
	CP072	Operation of personal transport equipment	No
	CP073	Transport services	No
CP08		*Communication*	
	CP081	Postal services	No
	CP082	Telephone and telefax equipment	Yes
	CP083	Telephone and telefax services	No
CP09		*Recreation and culture*	
	CP091	Audio-visual, photographic and information procession equipment	Yes
	CP092	Other major durables for recreation and culture	Yes
	CP093	Other recreational items and equipment, gardens and pets	Yes
	CP094	Recreational and cultural services	No
	CP095	Newspapers, books and stationary	Yes
	CP096	Package holidays	Yes
CP10		*Education*	No
CP11		*Restaurants and hotels*	
	CP111	Catering services	No
	CP112	Accommodation services	No
CP12		*Miscellaneous goods and services*	
	CP121	Personal care	No
	CP122	Prostitution	No
	CP123	Personal effects n.e.c.	No
	CP124	Social protection	No
	CP125	Insurance	No
	CP126	Financial services n.e.c.	No
	CP127	Other services	No

Table 2.4: Classification of COICOP categories as goods or services.

This classification is based on the information given by Eurostat. Two categories cannot be fully linked to either goods or services. For CP044 and CP072 consist of both categories at a three-digit classification level. However, household consumption data is only available at the given granularity and therefore does not allow for a more precise classification.

NACE Rv.2 classification

Table 2.5 lists all two-digit NACE categories and their classification as goods or services.

Services	Goods	Housing
Financial and insurance activities	Agriculture, forestry and fishing	Real estate activities
Construction	Industry (except construction)	
Scientific and technical activities; administrative and support service activities	Manufacturing	
Information and communication		
Public administration, defence, education, human health and social work activities		

Table 2.5: Classification of NACE categories as Goods, Services or Housing

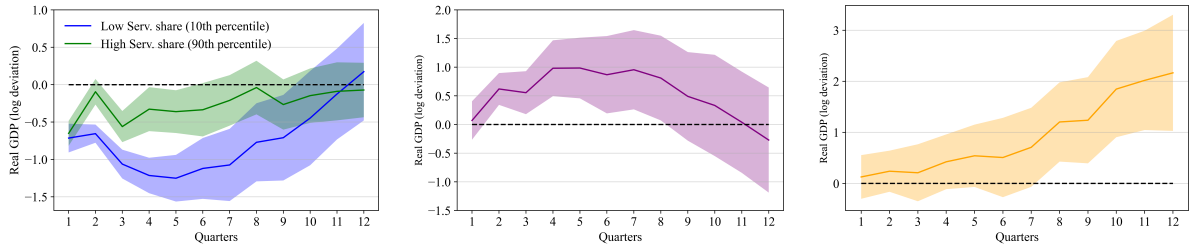
Empirical Extensions

We run multiple extensions of the baseline local projection to assess effects of additional channels at play. We highlight four main extensions: The role of income inequality, the response of tradable and service price indeces, the response of real exchange rates and the expenditure share on government services.

Income inequality may affect the response of real GDP through various channels - one of which is the non-homothetic preference channel presented above. Therefore, we augment the baseline LP with the average Gini coefficient on disposable income as follows:

$$y_{t+h,k} - y_{t-1,k} = \alpha + \beta^h i_t + \gamma^h (i_t * \bar{\omega}_k) + \lambda^h (i_t * \overline{Gini}_k) + \sum_{s=1}^{p} \Gamma_s^h y_{t-s,k} + \phi_k + u_{t+h,k}. \tag{2.27}$$

Figure 2.18 presents the IRFs for countries of different service consumption shares controlling for income inequality. The IRFs are constructed as the linear combination of β^h and γ^h with $Gini_c$ fixed at the sample mean. Controlling for income inequality does not alter the effect of the national consumption basket on output responses to monetary policy shocks. The share of service consumption seems to play the key role in driving differences in responses for the majority of the post-shock period. High income inequality is also associated with weaker responses to monetary policy shocks, however, significantly so only after nine quarters.



(a) Services share (b) Gamma Difference (c) Lambda Difference

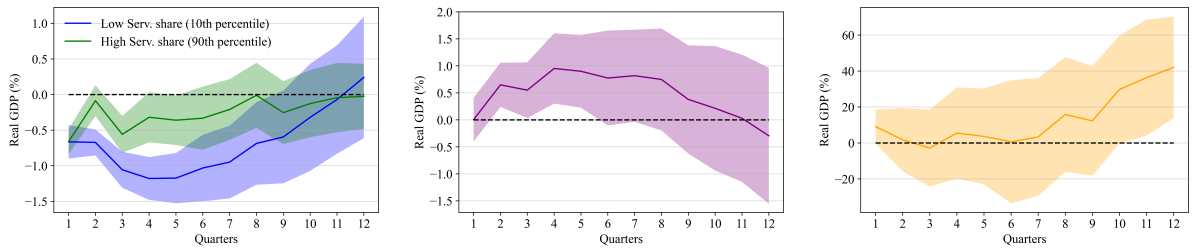
Figure 2.18: Impulse response functions of real GDP to a contractionary monetary policy shock of one standard deviation for high and low service consumption shares controlling for income inequality.

Notes: In graph a), the low service (S) share line is based on the 10th percentile of service shares in the sample and the high service share line on the 90th percentile keeping Gini fixed at the sample mean. Graph b) shows the interaction term coefficient γ. Graph c) shows the interaction term coefficient λ. The shaded areas represent 90th percent confidence intervals. Standard errors are clustered around countries.

The share of constrained Hand-to-Mouth households (HtM) across countries plays a major role in defining responses to shocks. While we control for these shares indirectly in our baseline estimation with country fixed effects, we extend our analysis to control for them in the local projection directly. We use the shares of HtM agents from Almgren et al. (2022). The resulting impulse response function show no significant difference to the baseline results. We augment the baseline regression by including the HtM shares as follows:

$$y_{t+h,c} - y_{t-1,c} = \alpha + \beta^h i_t + \gamma^h (i_t \times \bar{\omega}_c) + \lambda^h (i_t \times HtM_c) + \sum_{s=1}^{p} \Gamma_s^h y_{t-s,c} + \phi_k + u_{t+h,c}. \tag{2.28}$$

Figure 2.19 presents the IRFs for countries of different service consumption share controlling for the share of HtM agents. The IRFs are constructed as the linear combination of β^h and γ^h with HtM_c fixed at the sample mean. Controlling for HtM shares does not alter the effect of the national consumption basket on output responses to monetary policy shocks.



(a) Services share (b) Gamma Difference (c) Lambda Difference

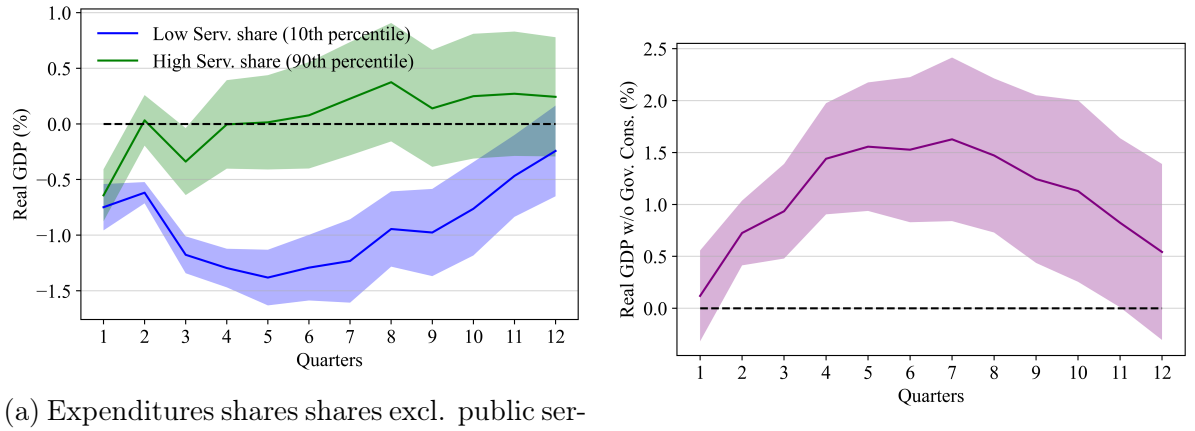
Figure 2.19: Impulse response functions of real GDP to a contractionary monetary policy shock of one standard deviation for high and low service consumption shares controlling for the share of HtM agents.

Notes: In graph a), the low service (S) share line is based on the 10th percentile of service shares in the sample and the high service share line on the 90th percentile keeping the HtM share at the sample mean. Graph b) shows the interaction term coefficient γ. Graph c) shows the interaction term coefficient λ. The shaded areas represent 90th percent confidence intervals. Standard errors are clustered around countries.

Government services are part of the service expenditure share in our baseline estimation. One potential reason countries with high service shares react less to a monetary policy shock could relate to a higher share of government services that react little to monetary policy. When excluding public services from the service expenditure shares, however, we do not find evidence for this channel as shown in Figure 2.20. The corresponding specification is as follows:

$$y_{t+h,c} - y_{t-1,c} = \alpha + \beta^h i_t + \gamma^h (i_t \times \tilde{\omega}_c) + \sum_{s=1}^{p} \Gamma_s^h y_{t-s,c} + \phi_k + u_{t+h,c}. \tag{2.29}$$

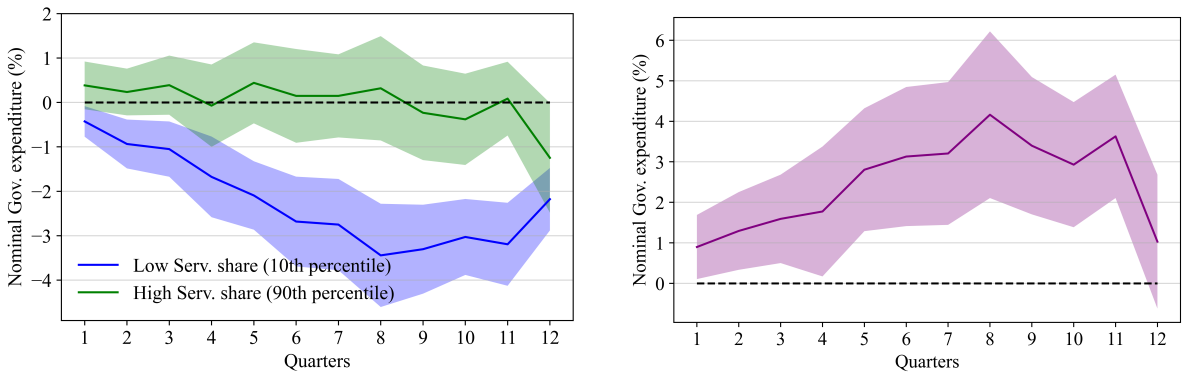
where $\tilde{\omega}_c$ corresponds to a county's share of service expenditure excluding the expenditure shares on government services.



(a) Expenditures shares shares excl. public services (b) Difference

Figure 2.20: Impulse response functions

Government expenditure responses are shown in Figure 2.21. shows the difference in responses of nominal government expenditure across countries with high and low service expenditure shares.



(a) High and low service share (b) Difference

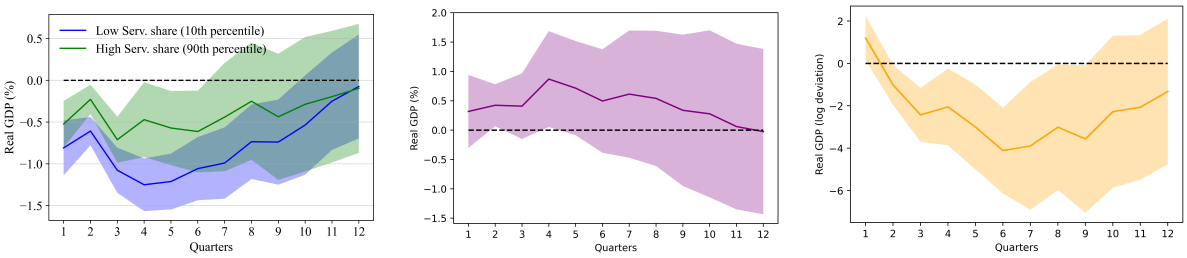
Figure 2.21: Impulse response functions of nominal government expenditure to a contractionary monetary policy shock of one standard deviation for high and low service consumption shares.

Notes: In graph a), the low service (NT) share line is based on the 10th percentile of service shares in the sample and the high service share line on the 90th percentile, respectively. Graph b) shows the interaction term coefficient γ. The shaded areas represent 90th percent confidence intervals. Standard errors are clustered around countries.

Non-durable vs durables. We run a regression controlling for durable goods. We divide the goods in durable and non-durable. We estimate the parameters of the following regression:

$$y_{t+h,c} - y_{t-1,c} = \alpha + \beta^h i_t + \gamma^h (i_t \times \bar{\omega}_c) + \lambda (i_t \times \bar{\omega}_{durable}) \sum_{s=1}^{p} \Gamma_s^h y_{t-s,c} + \phi_k + u_{t+h,c} \tag{2.30}$$

where we define $\omega_{durable}$ as the average 2000-2020 share of durable goods. The impulse response functions are plotted in Figure 2.22. High consumption of durable goods amplify the monetary policy shock. However, the sign and shape of the interaction terms between service expenditure share and monetary policy shock is almost unchanged.



(a) Non-durables vs durables (b) Gamma difference (c) Lambda difference

Figure 2.22: Non-durables vs durables

Relationship between service expenditure and housing expenditure shares When computing the service expenditure share ω we exclude housing consumption. In order to show that our results are not driven by difference in housing expenditure shares across countries we run the following regression:

$$\omega_{t,k} = \alpha + \gamma_t + \beta \sigma_{t,k} + \epsilon_{k,t}$$

where γ_t is a year fixed effects and σ corresponds to the housing expenditure shares for country k. Figure 2.23 shows the linear fit in a binned scatter plot. There is no significant relationship between our measure of service expenditure shares and housing expenditure shares.

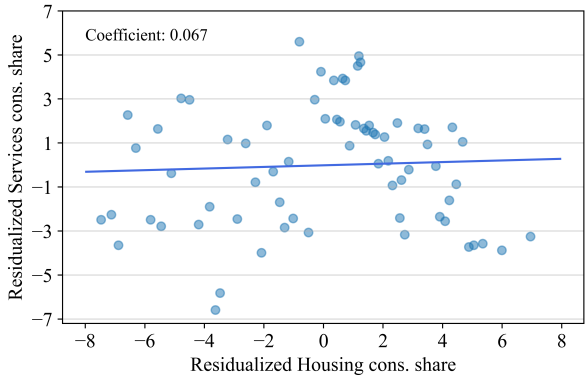


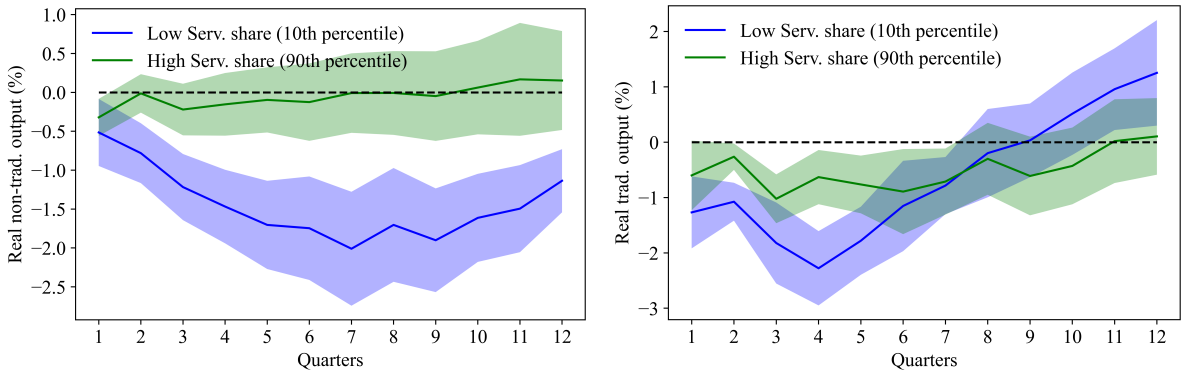
Figure 2.23: Services vs housing expenditure shares.

Output of goods and services. We estimate the following local projection:

$$y^{\iota}_{t+h,c} - y_{t-1,c} = \alpha + \beta^h i_t + \gamma^h (i_t \times \bar{\omega}_c) + \sum_{s=1}^{p} \Gamma^h_s y^{\iota}_{t-s,c} + \phi_k + u_{t+h,c} \tag{2.31}$$

where i_t is the monetary policy shock, ω_c is the 2000-2020 average service expenditure share, ϕ_k is a country fixed effect and $y^{\iota}_{t,c}$ can be the tradable ($\iota = T$) or service output ($\iota = N$).

Figure 2.24 shows the impulse responses function deriving from equation (2.31). The reaction of service output to a monetary policy shock depends substantially on service expenditure share. Countries with high-non tradable expenditure share service output and do not react much to a monetary policy shock, while service output contracts substantially in countries with low service expenditure share. Instead, tradable output reacts more similarly to a monetary policy shock in countries with different service expenditure share.

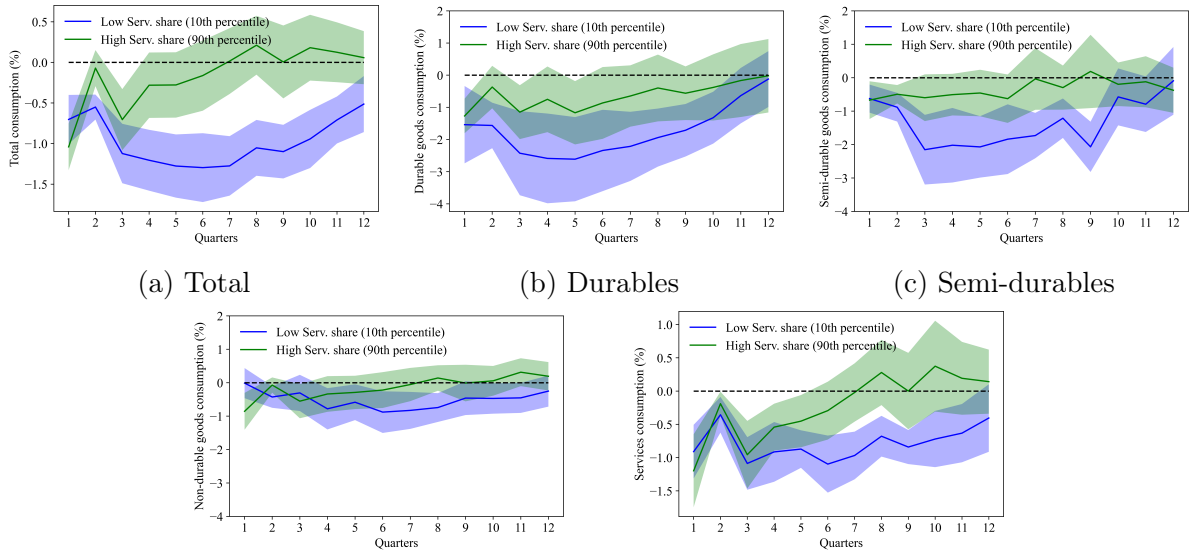


(a) Non-tradable output

(b) Tradable Output

Figure 2.24: Impulse response functions of goods and services output to a contractionary monetary policy shock of one standard deviation.

Consumption by sector. Figure 2.25 shows the difference in consumption responses across sectors for countries with high and low service expenditure shares. The graphs reveal two key findings. First, total consumption resembles output, with weaker responses in countries with higher service shares. Second, the difference in responses declines with the for less tradable sectors. While consumption in highly tradable, durable goods reacts significantly weaker for countries with larger service consumption shares, the difference is smaller for semi-durable goods and becomes insignificant for services.



(a) Total (b) Durables (c) Semi-durables

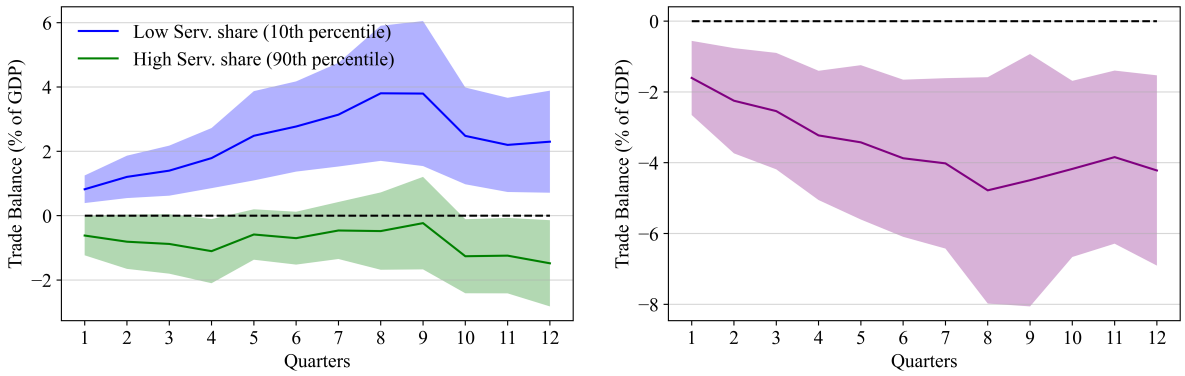
(d) Non-durables (e) Services

Figure 2.25: Impulse response functions of real consumption by sector to a contractionary monetary policy shock of one standard deviation.

The Trade Balance (% of GDP) reacts different for the two groups of countries as shown in Figure 2.26. We plot the cumulative impulse response of countries' trade balances to a contractionary monetary policy shock. We follow Ramey and Zubairy (2018) in computing the cumulative impulse response as follows:

$$\sum_{h=0}^{H}(y_{t+h,c} - y_{t-1,c}) = \alpha + \beta^h \sum_{h=0}^{H} i_{t+h} + \gamma^h (\sum_{h=0}^{H} i_{t+h} \times \bar{\omega}_c) + \sum_{s=1}^{p} \Gamma_s^h y_{t-s,c} + \phi_k + u_{t+h,c}.$$

The summation on the left-hand side corresponds to the cumulative difference of real GDP at every horizon h from the initial GDP level at time t. The coefficients $beta^h$ and $gamma^h$ are interpreted as the effect of the cumulative change of the monetary policy shocks interacted with the non-tradabable expenditure share $\bar{\omega}_c$. Countries with relatively low service consumption shares experience a larger increase in the trade balance than countries with high service consumption shares. A one standard deviation contractionary monetary policy shock results in a 0.5 percentage point increase in the trade balance after two quarters for countries with a service consumption share as low as the 10th percentile of service consumption shares in the sample.



(a) High and low service share

(b) Difference

Figure 2.26: Cumulative impulse response functions

Prices of Goods and Services are important determinants of responses to monetary policy shocks. In particular, we find that service prices react slower and weaker than tradable prices. This finding is in line with Cravino et al. (2020), highlighting the stronger nominal rigidities in service sectors. Figure 2.27 shows the differences in responses of service and tradable price indices to a contractionary monetary policy shock. The IRFs are constructed from the following local projection exercise

$$y_{t+h,c} - y_{t-1,c} = \alpha + \beta^h i_t + \sum_{s=1}^{p} \Gamma_s^h y_{t-s,c} + \phi_k + u_{t+h,c} \tag{2.32}$$

where the left-hand side is the deviation in the price index (2015=100) from its initial value. The graphs reveal that prices of goods respond more strongly to monetary policy shocks. While prices of goods adjust downwards immediately, service prices show a significant adjustment only after 4 and 8 quarters.[13]

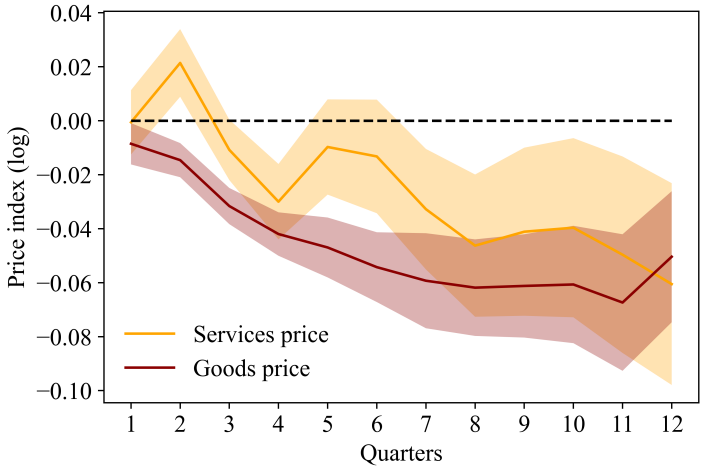


Figure 2.27: Impulse response of log goods, and service price indices.

Comparison of results with existing estimates

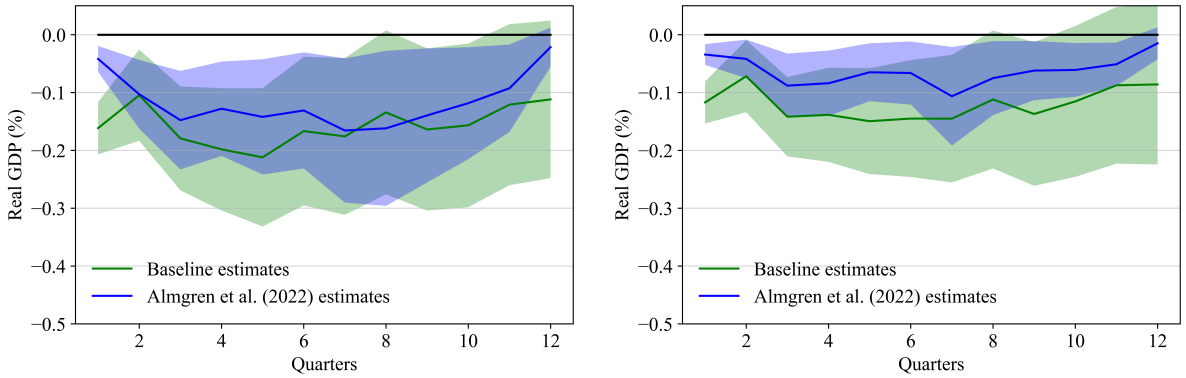
In order to check that our baseline results are in line with the estimates obtained in the existing literature we compare our estimates to Almgren et al. (2022). Specifically, we

[13]The price indices are computed based on the Nace Rev.2 price indices classified in Appendix 2.6.

run our baseline local projection for individual countries as follows:

$$y_{t+h,k} - y_{t-1,k} = \alpha + \beta^h i_t + \sum_{s=1}^{p} \Gamma_s^h y_{t-s,k} + u_{t+h,k}. \tag{2.33}$$

Figure 2.28 plots our estimates compared to theirs for Germany and the Netherlands. The impulse response functions to a one standard deviation monetary policy shock follow each other closely and are not distinguishable at the 90 percent confidence intervals.



(a) Germany

(b) Netherlands

Figure 2.28: Impulse response functions for individual countries.

2.7 Appendix - Model

Stationary Allocation

Consider a stationary allocation of the economy, where all variables are constant through time. Constant prices of individual commodities $P_t^S = P^S$ and $P_t^G = P^G$, jointly with arbitrage across the two types of bonds, imply that the interest rates on bonds denominated in units of tradable and nontradable must be equal: $R_t = R_t^m = R$. We can thus denote B_R as the aggregate wealth of the Ricardian household in the stationary allocation.

The stationary-allocation expenditure by the Ricardian household thus equals

$$e_R = W_R l_R + \Psi_R + (R-1)B_R \tag{2.34}$$

and for the Hand-to-Mouth:

$$e_H = W_H l_H + \Psi_H \tag{2.35}$$

From firms' demand for labor and allocation of profits across the two households, incomes of the two agents are given by $W_R l_R + \Psi_R = \xi_S y^S + \xi_G y^G$ and $W_H l_H + \Psi_H = (1-\xi_S)y^S + (1-\xi_G)y^G$. Hence, expenditure of the two households writes as:

$$e_R = \xi_S y^S + \xi_G y^G + (R-1)B_R, \text{ and } e_H = (1-\xi_S)y^S + (1-\xi_G)y^G. \tag{2.36}$$

The trade balance of the economy in steady state is equal to $TB = (R-1)B_R$. The households' allocation of expenditure across tradables and nontradables is also constant and governed by the conditions in (2.9) and (2.8). The aggregate demand for nontradables is given by the sum of the two households' demand, as follows:

$$y^S = c_H^S + c_R^S = e_H \omega_H^S + e_R \omega_R^S \tag{2.37}$$

which yields:

$$y^S = (y^G - TB)\frac{P^G}{P^S}\frac{\left[\omega_H^S\left(1-\xi_G\right)+\omega_R^S\left(\xi_G\right)\right]}{\left[1-\omega_H^S\left(1-\xi_S\right)-\omega_R^S\left(\xi_S\right)\right]}. \tag{2.38}$$

The relative allocation of labor of the two households across the two sectors satisfies:

$$\frac{l_H^S}{l_H^G} = \frac{\alpha_N}{\alpha_T}\frac{1-\xi_S}{1-\xi_G} \equiv x_H^S\frac{P^S y^S}{P^G y^G} \text{ and } \frac{l_R^S}{l_R^G} = \frac{\alpha_N}{\alpha_T}\frac{\xi_S}{\xi_G}\frac{P^S y^S}{P^G y^G} \equiv x_R^S. \tag{2.39}$$

Hence, labor input in the two sectors satisfies

$$l_H^S = \frac{x_H^S}{1+x_H^S}l_H, \text{ and } l_R^S = \frac{x_R^S}{1+x_R^S}l_R, l_H^G = \frac{1}{1+x_H^S}l_H, \text{ and } l_R^G = \frac{1}{1+x_R^S}l_R. \tag{2.40}$$

Pro-Cyclical Relative Value of Services Output

We show in this appendix that the relative value of service output is procyclical when preferences are non-homothetic, in a homogeneous monetary union. To ths purpose, it is convenient to express this relative value as function of the two households' income shares in the two sectors, and of their expenditure share on goods. To do so, note that total G-output can be written as, omitting time-subscripts:

$$P^G y^G = P^G c_R^G + P^G c_H^G = \omega_R^G e_R + \omega_H^G e_H.$$

In turn, expenditure of the two households writes as:

$$e_R = \xi_S P^S y^S + \xi_G P^G y^G \text{ and } \left(1-\xi_S\right)P^S y^S + \left(1-\xi_G\right)P^G y^G.$$

Hence, the relative value of output of the two sectors can be expressed as:

$$\frac{P^S y^S}{P^G y^G} = \frac{1-\omega_R^G\xi_G-\omega_H^G\left(1-\xi_G\right)}{\omega_R^G\xi_S+\omega_H^G\left(1-\xi_S\right)}.$$

When a contractionary monetary policy shock reduces aggregate output, both households have lower income and expenditure, and allocate a greater share of their expenditure to goods, i.e. $\omega_{j,1}^G > \omega_{j,2}^G$ for $j \in \{H, R\}$. As the expression for the relative value of output is decreasing in the two expenditure shares for goods, an aggregate contraction leads to a reduction in the relative value of service output, and vice-versa.

Effects of Monetary Policy on Total Output, Price of Goods, and Services Output

First, we show that there is a positive monotonic relationship between the output and the price of goods in the initial period $t = 1$ of this model economy, in the homogeneous monetary union. Hence, a mild decline in goods output is associated with a mild decline in their price. To see this, note that the relative labor demand in the two sectors can be combined with the technology in the G sector to yield an equation characterizing the supply of G as a function of its nominal price and of the rigid nominal wage:

$$P_1^G = \frac{\bar{W}^G}{\alpha_T}\left(y_1^G\right)^{\frac{1-\alpha_T}{\alpha_T}} \tag{2.41}$$

Second, consider the impact of monetary policy on output of services. Under Cobb-Douglas preferences, the fall in S-output is proportional to the fall in the value of G-output and given by:

$$y_1^S = \frac{1-\nu}{\nu}\frac{1}{P_1^S}\left(y_1^G\right)^{\frac{1}{\alpha^G}},$$

where $\left(y_1^G\right)^{\frac{1}{\alpha^G}} = P_1^G y_1^G$, and where P_1^S is equal to the stationary-allocation value of the price of services, due to the complete wage rigidity and unitary elasticity of output to labor input in the S-sector.

When preferences are non-homothetic, the impact of monetary policy on S-output is more than proportional to the decline of G output, as all households reallocate their expenditure away from luxuries when output contracts:

$$y_1^N = \zeta(P_1^G y_1^G, y_1^N)\frac{P_1^G y_1^G}{P_1^S}, \tag{2.42}$$

where $\zeta(P_1^G Y_1^G, Y_1^N) = \frac{P_{N,1}y_{N,1}}{P_{T,1}y_{T,1}}$ is a shifter that reflects the country's preference for services. This shifter is procyclical under the conditions stated in Proposition 7, as Appendix 2.7 shows. Thus, services output moves in lockstep with tradable output.

Finally, the effects of monetary policy on total output are simply equal to the sum of the effects on output of the two sectors. If the distribution of income reduces the magnitude of the effects of monetary policy on output of tradable good, it also reduces the effects on output of good S and therefore on total output.

Conditions for Positive Relationship between Price and Consumption of Goods in Heterogeneous Monetary Union

First, we express Ricardian consumption of goods at $t = 1$ as:

$$c_{R,1}^G = \left(\xi_G y_1^G + \xi_S \frac{P_t^S}{P_t^G} y_1^S - b_{R,2}\right)^{(1-\epsilon)} (P_1^G)^{-\epsilon}. \tag{2.43}$$

Second, we impose for simplicity that $\frac{P_t^S}{P_t^G} y_1^S = \frac{1-\nu}{\nu} c_t^G$, as would be the case under Cobb-Douglas preferences. With non-homothetic preferences, services output also depends on an extra wedge that follows from the aggregation of non-homothetic demand for services. This wedge is however increasing in total consumption, as services are a luxury. Hence, this simplifying assumption understates the magnitude of the increasing relationship between services output and aggregate goods consumption. Hence, if we can prove that $c_{R,1}^G$ is increasing in P_1^G, the result will be true a fortiori under non-homothetic preferences.

Third, we can use the fact that total goods consumption in country i equals the sum of total output of goods in the same country and the savings/borrowing of the Ricardian household, $c_1^G = y_1^G - b_{R,2}$. Thus, we can write:

$$c_{R,1}^G = \left(\left(\xi_G + \xi_S \frac{1-\nu}{\nu}\right) y_1^G - \left(1 + \xi_S \frac{1-\nu}{\nu}\right) b_{R,2}\right)^{(1-\epsilon)} (P_1^G)^{-\epsilon}. \tag{2.44}$$

Finally, we can use (2.41) to write the output of goods as a function of the prices of goods:

$$y_1^G = \left(\frac{\alpha_T P_1^G}{W_1^G}\right)^{\frac{\alpha_T}{1-\alpha_T}}, \tag{2.45}$$

and plug in (2.44):

$$c_{R,1}^G = \left(\left(\xi_G + \xi_S \frac{1-\nu}{\nu}\right)\left(\frac{\alpha_T P_1^G}{W_1^G}\right)^{\frac{\alpha_T}{1-\alpha_T}} - \left(1 + \xi_S \frac{1-\nu}{\nu}\right) b_{R,2}\right)^{(1-\epsilon)} (P_1^G)^{-\epsilon}. \tag{2.46}$$

Ricardian consumption of goods is a function of the price of goods: if the price of goods increases, output of goods increases too. Higher prices of goods also increase consumption and output of services. Thus, the income and consumption of Ricardian household would increase, too. However, as expenditure increases, the Ricardian household reduces the goods' share of its expenditure, in additional to the conventional substitution effect. Therefore, higher goods' prices can also reduce goods' consumption. Differentiating $c^G_{R,1}$ with respect to P^G_1, for a given value of $b_{R,2}$ we obtain:

$$\frac{dc^G_{R,1}}{dP^G_1} = \frac{(1-\epsilon)\frac{\alpha_T}{1-\alpha_T}\left(\xi_G + \xi_S\frac{1-\nu}{\nu}\right)\frac{y^G_1}{P^G_1}}{\left(\xi_G + \xi_S\frac{1-\nu}{\nu}\right)y^G_1 - \left(1+\xi_S\frac{1-\nu}{\nu}\right)b_{R,2}} - \frac{\epsilon}{P^G_1} > 0. \tag{2.47}$$

From this derivative, we obtain a condition on the ratio between Ricardian savings or borrowings and output of goods that is consistent with a positive overall effect of an increase in the price of goods on the Ricardian consumption of goods:

$$-\frac{b_{R,2}}{y^G_1} > \frac{\alpha-\epsilon}{\epsilon(1-\alpha)}\frac{\xi_G + \xi_S\frac{1-\nu}{\nu}}{1+\xi_S\frac{1-\nu}{\nu}}. \tag{2.48}$$

If $\alpha > \epsilon$ and for ϵ small enough $\frac{\alpha-\epsilon}{\epsilon(1-\alpha)}\frac{\xi_G+\xi_S\frac{1-\nu}{\nu}}{1+\xi_S\frac{1-\nu}{\nu}}$ is a sizable number. We set parameters so that $\frac{\alpha-\epsilon}{\epsilon(1-\alpha)}\frac{\xi_G+\xi_S\frac{1-\nu}{\nu}}{1+\xi_S\frac{1-\nu}{\nu}} > 1$. The above condition is thus satisfied as long as the monetary policy shock generates external balances that are smaller than output.

Proofs for Propositions and Lemmas

This appendix provides more formal proofs for the propositions and lemmas presented in Section 2.4 of the main text. We refer to the model equations as defined in Section 2.3. Assumptions used are Assumption 2.4 ($\varepsilon > 0$) and Assumption 2.4 (e.g., $\xi_{S,R,k} > \xi_{G,R,k}$ where relevant for pro-cyclical inequality).

Proofs for Section 2.4.1 (Expenditure Shares)

Proof of Lemma 3. The expenditure share on services for household j in country k is given by equation (2.9):

$$\omega^S_{j,k}(e_{j,k}) = 1 - \nu\left(\frac{P^S_k}{e_{j,k}}\right)^{\varepsilon}\left(\frac{P^G}{P^S_k}\right)^{\gamma}.$$

Differentiating $\omega^S_{j,k}$ with respect to $e_{j,k}$, holding prices P^S_k and P^G constant:

$$\begin{aligned}\frac{\partial\omega^S_{j,k}}{\partial e_{j,k}} &= -\nu(P^S_k)^{\varepsilon-\gamma}(P^G)^{\gamma}\frac{\partial}{\partial e_{j,k}}(e^{-\varepsilon}_{j,k}) \\ &= -\nu(P^S_k)^{\varepsilon-\gamma}(P^G)^{\gamma}(-\varepsilon e^{-\varepsilon-1}_{j,k}) \\ &= \varepsilon\nu(P^S_k)^{\varepsilon-\gamma}(P^G)^{\gamma}e^{-\varepsilon-1}_{j,k}.\end{aligned}$$

Given Assumption 2.4 ($\varepsilon > 0$), and that prices and expenditure are positive, $P^S_k > 0$, $P^G > 0$, $e_{j,k} > 0$, it follows that $\frac{\partial\omega^S_{j,k}}{\partial e_{j,k}} > 0$. □

Proof of Lemma 4. The average service expenditure share is $\bar{\omega}^S_k(\bar{e}_k) = x_{H,k}\omega^S_{H,k}(x_{H,k}\bar{e}_k) + x_{R,k}\omega^S_{R,k}(x_{R,k}\bar{e}_k)$, where $x_{j,k} = e_{j,k}/\bar{e}_k$ is the share of household j's expenditure in total

country expenditure, held constant for this partial derivative. Differentiating $\bar{\omega}_k^S$ with respect to $\bar{e}_k$:

$$\frac{\partial \bar{\omega}_k^S}{\partial \bar{e}_k} = x_{H,k} \frac{\partial \omega_{H,k}^S(e_{H,k})}{\partial e_{H,k}} \frac{\partial e_{H,k}}{\partial \bar{e}_k} + x_{R,k} \frac{\partial \omega_{R,k}^S(e_{R,k})}{\partial e_{R,k}} \frac{\partial e_{R,k}}{\partial \bar{e}_k}.$$

Since $e_{j,k} = x_{j,k}\bar{e}_k$, then $\frac{\partial e_{j,k}}{\partial \bar{e}_k} = x_{j,k}$. Substituting this in:

$$\frac{\partial \bar{\omega}_k^S}{\partial \bar{e}_k} = x_{H,k}^2 \frac{\partial \omega_{H,k}^S(e_{H,k})}{\partial e_{H,k}} + x_{R,k}^2 \frac{\partial \omega_{R,k}^S(e_{R,k})}{\partial e_{R,k}}.$$

From Lemma 3, $\frac{\partial \omega_{j,k}^S(e_{j,k})}{\partial e_{j,k}} > 0$ under Assumption 2.4 ($\varepsilon > 0$). Since $x_{j,k}^2 \geq 0$ (and typically > 0), the sum of positive terms is positive. Thus, $\frac{\partial \bar{\omega}_k^S}{\partial \bar{e}_k} > 0$. □

Proof of Lemma 5. As stated in the main text, the average service expenditure share is given by:

$$\bar{\omega}_k^S = 1 - \nu \left(\frac{P_k^S}{\bar{e}_k}\right)^{\varepsilon} \left(\frac{P^T}{P_k^S}\right)^{\gamma} \left(x_{H,k}^{1-\varepsilon} + x_{R,k}^{1-\varepsilon}\right).$$

Let $K = \nu \left(\frac{P_k^S}{\bar{e}_k}\right)^{\varepsilon} \left(\frac{P^T}{P_k^S}\right)^{\gamma}$. Since $\nu > 0$ and prices and aggregate expenditure are positive, $K > 0$. Then $\bar{\omega}_k^S = 1 - K\left(x_{H,k}^{1-\varepsilon} + x_{R,k}^{1-\varepsilon}\right)$. Consider the function $f(x) = x^{1-\varepsilon}$. Its second derivative is $f''(x) = (1-\varepsilon)(-\varepsilon)x^{-\varepsilon-1} = -\varepsilon(1-\varepsilon)x^{-\varepsilon-1}$. Under Assumption 2.4 ($\varepsilon > 0$), and assuming $0 < \varepsilon < 1$ (as typical for Boppart (2014) preferences) then $1 - \varepsilon > 0$. Thus, $f''(x) < 0$, meaning $f(x) = x^{1-\varepsilon}$ is a strictly concave function. By Jensen's inequality, for a concave function f, a mean-preserving spread in its arguments decreases the expected value of the function's evaluations. Therefore, an increase in the dispersion $\sigma_{x,k}$ of $(x_{H,k}, x_{R,k})$ (which is a mean-preserving spread since $x_{H,k} + x_{R,k} = 1$) will decrease the sum $(x_{H,k}^{1-\varepsilon} + x_{R,k}^{1-\varepsilon})$. Since $\bar{\omega}_k^S = 1 - K \times$ (term that decreases with dispersion), and $K > 0$, it follows that $\bar{\omega}_k^S$ increases with the dispersion $\sigma_{x,k}$. □

Chapter 3

Economic Consequences of Large Extraction Declines: Lessons for the Green Transition

Rudolfs Bems
IMF and CEPR

Lukas Boehnert
University of Oxford

Andrea Pescatori
IMF

Martin Stuermer
IMF

Email: rbems@imf.org, lukas.boehnert@economics.ox.ac.uk, apescatori@imf.org, mstuermer@imf.org.

We are grateful for comments by Lukas Boer, Pierre-Olivier Gourinchas, Francesco Grigoli, Petya Koeva-Brooks, Rick van der Ploeg and Antonio Spilimbergo as well as participants of the IMF Research Department's Commodities Unit Seminar, the Centre for Study of African Economies Seminar, the Ruhr Graduate School Symposium, the Summer Meeting of the North American Econometric Society in Los Angeles, and the Summer Meeting of the Central Bank Research Association Meeting in New York City. Abdullah Albijadi, Rachel Brasier, Wenchuan Dong, Joseph Moussa, Abreshmi Nowar, and Tianchu Qi provided outstanding research assistance. The views in this paper are those of the authors and do not necessarily reflect the views of the International Monetary Fund, its Executive Board, or its management.

Abstract

Limiting climate change requires an 80 percent reduction in fossil fuel extraction until 2050. What are the macroeconomic consequences for fossil fuel producing countries? We identify 35 episodes of persistent, exogenous declines in extraction based on a new dataset for 13 minerals (oil, gas, coal, metals) and 122 countries since 1950. We use local projections to estimate effects on real output as well as the external and the domestic sectors. Declines in extractive activity lead to persistent negative effects on real GDP and the trade balance. The real exchange rate depreciates but not enough to offset the decline in net exports. Effects on low-income countries are significantly larger than on high-income countries. Results suggest that legacy effects of bad institutions could prevent countries from benefiting from lower resource extraction.

3.1 Introduction

Reaching net zero carbon emissions by 2050 requires an 80 percent reduction in global fossil fuel extraction compared with 2021 levels, according to the International Energy Agency (2022). Many countries could therefore experience a permanent contraction in their fossil fuel extractive sectors. What would be the macroeconomic impact?

A large body of literature emphasizes the negative impact of having a sizeable extraction sector on a country's economic growth, i.e., the resource curse, by weighing on the performance of manufacturing (see e.g., Frankel (2012) and Krugman (1987)) and on the quality of institutions (see Mauro (1995) and Lane and Tornell (1996)).

There is, however, a dearth of analyses on the macroeconomic effects of a reversal. It is an open question whether a decline in fossil production is detrimental or beneficial for countries' economic growth. Can the maladies identified by the resource curse literature during resource booms be reversed by shrinking the extractive sector? As climate policies and their effects on a specific country's extraction sector are uncertain, to what extent does anticipation or the lack thereof shape the effects of an extraction decline on economic activity, especially on the trade balance?

This paper sheds light on these questions by estimating the macroeconomic impact of large and persistent, exogenous declines in extractive activity. We study major transmission channels, the role of institutions, and compare the empirical results to a neoclassical model with and without full anticipation.

We collect a new data-set on the extraction of 13 commodities, including crude oil, coal, and metals, for countries worldwide starting in 1950. Using statistical and narrative approaches, we identify 35 episodes of persistent declines in extractive activity that are not correlated with other economic shocks hitting the economy at the same time. We set up a stylized classical small-open-economy real model to study adjustments to the extraction declines with and without anticipation. We use local projections to estimate the effects of these episodes on real output, the exchange rate, the quality of institutions, and the external and the domestic sectors.

The typical episode in our new data-set is a large and persistent decline in extractive activity: a 10 percent initial decline that accumulates to a 40 percent decline in the tenth year. Local projection methods show that such a decline leads to a persistent decline in real GPD of six percent after about six years.

Both private and public consumption as well as investment fall in line with the decline in GDP. Consumption and the exchange rate, however, have a delayed reaction, inconsistent with full anticipation of the persistent fall in extraction and related revenues. The real exchange rate depreciates, eventually by about 20 percent, but not enough to stimulate a reallocation of economic activity towards other trade-able sectors such as manufacturing that could potentially offset the decline in exports. Empirical results, in fact, reveal significant negative spillover effects onto both the manufacturing and services sectors. Net exports fall in line with extraction.

The analysis finds heterogeneity of the effects across countries. This heterogeneity depends on their per capita income and the size of the mining or extractive sector[1] relative to the economy. A large and persistent decline in extraction implies significantly larger declines in economic activity for low-income countries than for high-income countries.

Our results show that a decline in extractive activity does not improve the quality of institutions not even a decade after the shock. Theory stresses that increases in resource extraction diminishes long-term growth through rapacious rent-seeking (the voracity effect), negatively affecting the quality of institutions through political economy effects (as in Lane and Tornell (1996) and Tornell and Lane (1999)) and by increasing corruption (see Mauro (1995) and Leite and Weidmann (1999)). Our contribution to this literature is to show that there is no reversion of the deterioration of institutions after a decline in resource extraction. This suggests hysteresis effects and an asymmetric response of institutions to shocks: once institutions are damaged it is difficult to improve them. Stressing the relevance of good institutions, results also show that starting with higher-quality institutions can help to buffer the negative economic effects of persistent declines in extractive activity.

Our results also uncover negative spillover effects from the extractive sector to the manufacturing and service sectors. The literature stresses that high resource extraction makes countries susceptible to the Dutch Disease. The real exchange rate becomes overly appreciated in response to positive commodity price shocks, causing a contraction of the tradable sector, especially manufacturing (e.g., Frankel (2012) and Krugman (1987)). We are the first to show that for the reverse case of decreasing resource extraction, the

[1]We use mining and extractive sector interchangeably. Following the North American Industry Classification System, the mining sector is defined as comprising all establishments that extract naturally occurring mineral solids, such as coal and ores; liquid minerals, such as crude petroleum; and gases, such as natural gas. We use the term mining sector interchangeably with extractive sector.

depreciation of the exchange rate does not stimulate a reallocation towards other tradable sectors such as the manufacturing sector, pointing possibly towards long lasting scars from resource boom. Since these sectors provide inputs and process outputs of the extractive sector, they are also negatively affected to an extent that more than offsets the potential benefits of the depreciation of the real exchange rate.

The initial share of the manufacturing sector in value added matters, though. Economies with a bigger initial manufacturing share fare better. This suggests that there are non-convexities in the tradable sector (i.e., sunk costs), favoring the intensive margin (existing exporting manufacturing firms) rather than the extensive margin (the establishments of new manufacturing firms).

We find little anticipation effects for extraction declines. This is in contrast to extraction booms, which are anticipated by large mineral discoveries and lead to an initial domestic saving-investment imbalance followed by the export boom as found in Arezki et al. (2017). It is, indeed, surprising how well a stylized classical small-open-economy real model can match qualitatively the empirical results once it departs from full anticipation. This is particularly the case, when the extraction decline is perceived as a series of temporary negative shocks that lead expectations of mining output to adjust slowly over time.

The paper runs a battery of robustness checks. Our results are robust to different specifications, the inclusion of various control variables and to using the synthetic control methodology. We also check for sensitivity to sub-samples and anticipation effects.

The main implication for policy makers is that the transition towards clean energy might be a challenge, especially for countries that rely on fossil fuel exports and are low to middle income. Countries will need to improve the quality of their institutions and policymakers should reduce climate policy uncertainty which may lead to a lack of anticipation, risking a more expensive adjustment process.

Our paper is to our knowledge the first to show that there are substantial impacts of declines in extractive activities at the macroeconomic level. It contributes to a broad literature on the resource curse (e.g., Sachs and Warner (2001) and Sala-i-Martin and Subramanian (2008)), as summarized in Brunnschweiler and Bulte (2008) and Van der Ploeg and Venables (2012) that focuses on the macroeconomic effects of resource booms. This literature supports the theoretical prediction that higher resource extraction dam-

ages long-term growth (see e.g., Sachs and Warner (2001), Feyrer et al. (2017), Allcott and Keniston (2018), Arezki et al. (2017), Cust et al. (2019), and Harding et al. (2020)). There is a string of the literature that studies periods of resource busts, which is, however, focused on regional economic impacts (see e.g., Black et al. (2005), Cavalcanti et al. (2019), B. Watson et al. (2023), Jacobsen and Parker (2016), and Hanson (2023)).

Finally, our results can provide empirical benchmarks to an emerging literature on the impact of the clean energy transition on fossil fuel producing countries. This literature has mostly focused on the economic implications of stranded assets (see e.g., Van der Ploeg and Rezai (2020)) and fiscal implications (see e.g., Mirzoev et al. (2020)) so far.

The paper is structured as follows. Section 3.2 introduces our new data-set and explains how we identify the shock episodes. Section 3.3 presents the macroeconomic effects of these shock episodes based on simple plots. Section 3.4 introduces our theoretical framework and its results. Section 3.5 explains the empirical framework and Section 3.6 present the empirical results. Section 3.7 examines the relationship between persistent extraction declines and the quality of institutions. Section 3.8 provides robustness checks and Section 3.9 concludes.

3.2 Identification Strategy: Episodes of Exogenous Extraction Declines

The main concern when estimating the causal effect of a decline in extractive activity on macroeconomic variables is reverse causality and omitted variables. In particular, events that affect mining production may equally be correlated to GDP (or other variables) or, to put it differently, a decline in economic activity may have repercussions on mining output in the same year. To address this substantial issue, we build a new data-set consisting of fossil fuels and mineral extraction, mineral prices, and macroeconomic variables; then starting from a far larger set of candidate episodes, we identify 35, macro-economically relevant, exogenous episodes of declines in extraction output for 11 commodities across 30 countries using a narrative approach. These 35 episodes are plausibly unrelated to other development in the domestic economy or global activity and can, thus, be used to infer their causal effect on the country's macroeconomic variables.

3.2.1 The Dataset

We assemble data on mining (or primary) production of fossil fuels and metals, and identify episodes of sustained declines. We collect an unbalanced data-set for the 13 largest commodities in terms of their worldwide production value for 122 countries from 1950 to 2020. Please find the list of commodities and descriptions in Section 3.10.

We combine annual data on the extracted quantities from different sources. We take historic primary production data for 11 commodities from Mitchell (2013) and include data on copper and lead production from the Federal Institute for Geosciences and Natural Resources (2012). The global historic production of metals is taken from U.S. Geological Survey (2012).

We take historical world prices of coal and metals from Schmitz (1979) and Schwerhoff and Stuermer (2019), and prices for oil and natural gas from BP (2022). We account for inflation by using the US consumer price index to deflate the commodity price data. We collect country-level data on macroeconomic variables including real GDP in local currency, nominal GDP, investment, private and public consumption, exports and imports, the real exchange rate (relative to the US Dollar), the current account and employment. We source the data from Feenstra et al. (2015). We use data on real value added by industry from the United Nations (2022). Finally, we include measures on political stability and institutional quality from the Center for Systemic Peace (2022) following Acemoglu et al. (2019).

3.2.2 Large, Persistent, and Macro-relevant Extraction Declines

A candidate episode is any extraction pattern $\{Q_{t+k}\}_{k=0}^{T}$ over a given time window, $[t, T]$, for each country and commodity. To obtain a measure of its macroeconomic importance for each identified episode i, we compute the *average value of the extraction decline (ADV)* relative to the country's nominal GDP at the start of the episode:

$$ADV_i = \frac{1}{T_i} \frac{\sum_{k=1}^{T_i} \Delta Q_{t+k,i} P_{t,i}}{GDP_{t,i}}, \tag{3.1}$$

where Q and P are the production and price of the commodity considered and T is the duration of the episode.[2] The duration of an episode is defined as the number of years

[2]The ADV is closely related to the Net Present Value (NPV), which is used in empirical work on resource windfalls like in Arezki et al. (2017). We can not compute the NPV, because the duration of the exogenous declines is ex-ante uncertain and the discount factor could vary across countries and

between the production peak (t=0) to the trough (t=T) where the trough is identified as the first year in which the fossil fuel sector operates at the new, permanent production level after the decline.

Episodes are initially selected using three criteria pertaining to the size, persistence, and macro-relevance of the decline in line with our objective to learn more about the potential impact of the energy transition: (a) The cumulative decline in the quantity of production is at least 20 percent from peak to trough; (b) after reaching the trough, the production level has to remain below its pre-shock level for at least 5 years; (c) the ADV is at least 0.05 percent. Criteria (a) and (b) isolate 154 episode, adding criterion (c) reduces that number to 70 macro-relevant episodes.

The value of the declines can be sizeable. The largest declines in terms of average revenue losses took place in oil, copper and gold production. Especially declines in crude oil production seem to come with large revenue losses of almost 20 percent of GDP on average (Section 3.10 shows the average size and duration).

3.2.3 Exogenous Extraction Declines

To identify episodes of extraction declines that are plausibly exogenous, we categorize them first by six different causes based on a narrative historical analysis (in the spirit of C. D. Romer and D. H. Romer (2010) and Ramey (2011) among others):

a. structural transitions (e.g. the breakdown of the Soviet Union) or wars,

b. global recessions,

c. policy changes affecting all domestic industrial sectors,

d. policy changes affecting only the extractive sector,

e. depletion,

f. undefined causes.

Episodes in categories (d.) and (e.) are classified as exogenous since the driving factors are supposedly exogenous to (aggregate) domestic economic conditions and do not affect other sectors at the same time. In such cases, the most typical decline is either the result of a change in domestic policy specific to the commodity sector or depletion.

Domestic policy changes include the nationalization of mines or changes to the tax

time.

regime. For example, Guyana nationalized its bauxite mines in 1975. Subsequent mismanagement drove down their efficiency and resulted in large declines in production. An example for a tax change, is the sudden tax increase on bauxite mining in Suriname in 1974. This tax increase only affected bauxite mining but not other sectors. The levy led to a decline in bauxite mining by more than 30 percent over the next five years and a permanently lower production level thereafter. We exclude episodes during which policy changes are part of a broader change in economic policies (e.g., the nationalization of all mining industries and redistribution of land in Zimbabwe in 1999).

Exogenous production declines due to depletion encompass instances, when the quality of the deposit or reservoir deteriorates for geological reasons or due to missing investment. These two factors often interact with each other. In any case, the result is higher production costs, making production unprofitable and forcing the mine to close down. One example, is iron ore mining in Eswatini in 1975, where lower ore quality made mining unprofitable and resulted in fast declines in production and the subsequent closure of mines. Another example is Chad, where aging oil wells and declining reservoir quality led to higher production costs and made investment unprofitable after 2005.

	1 # of Decline Episodes	2 Decline of Extraction Quant. (% of Peak)	3 Decline of Extraction Value (Mil. USD)	4 Duration (Years)	5 Annualized Decline Value (% of GDP)
Bauxite	3	-50	-96	5	-1.9
Copper	3	-78	-619	13	-0.6
Crude Oil	18	-45	-10,000	9	-2.5
Gold	3	-52	-2,060	7	-0.4
Lead	1	-78	-11	6	-0.1
Natural Gas	2	-54	-4,480	6	-0.1
Nickel	1	-83	-98	7	-0.1
Iron	1	-78	-32	5	-1.7
Silver	1	-44	-14	8	-0.9
Tin	1	-65	-236	6	-1.2
Zinc	1	-99	-95	7	-1.1
Total	35				

Table 3.1: Summary statistics of the exogenous declines in extractive activity by commodity.
Notes: Column 5 (the average ADV) is the result of dividing column 3 by column 4 and the GDP at time $t = 0$ for every shock episodes.

Focusing only on the exogenous ones, leaves us with 35 episodes of large and persistent declines. Table 3.1 provides the summary statistics, while Section 3.10 lists all episodes with more details about their narrative.

Figure 3.1 shows the distribution of these declines over time. The occurrence of shock episodes is relatively evenly distributed and is unrelated to the global economic cycle.

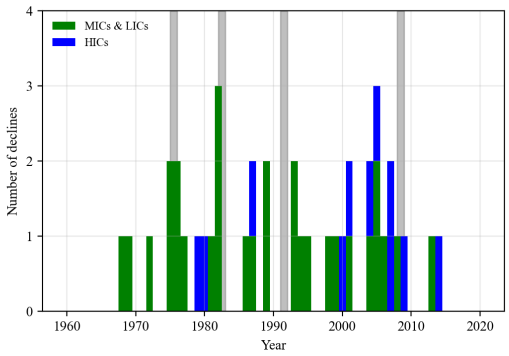


Figure 3.1: Distribution of extraction decline episodes over time.
Notes: MIC = Middle Income Countries, LIC = Low Income Countries, HIC = High Income Countries. The shaded areas represent global recessions as defined by Kose et al. (2020).

3.2.4 Anticipation

Declining extractive activity due to lower ore quality but also due to policy changes could be anticipated, leading the economy to start the adjustment beforehand. Anticipation would bias our estimation results towards a smaller negative impact on aggregate output, because the regression would not capture the initial adjustment.

To assess the extent to which shocks were anticipated, we reviewed forecasts of commodity production in Article IV consultation reports from the International Monetary Fund (IMF) and compared them to the actual extraction declines. Section 3.10 describes our methodology in detail.

The IMF reports aim at summarizing all relevant macroeconomic information based on discussions and information provided by the respective country's authorities, the private sector and IMF staff estimates. The biannual reports also typically provide a 5-year outlook for the economy and give guidance on macroeconomic and fiscal policies. In doing so, we assume that authorities interpret anticipated production declines as anticipated revenue declines.

We find evidence that the initial decline in production was only anticipated in 4 out of 28 decline episodes with Article IV Staff Reports (see Section 3.10). There was no information available for 7 of the 35 episodes in our sample. We keep the episodes that were anticipated in the sample but check the sensitivity of our results to anticipation in Section 3.8.2.

We also conduct a detailed examination of revisions in commodity volume projections for the entire episodes for Chad and Gabon. We choose these countries, because they

feature subsequent IMF Article IV consultation reports that cover most of the episodes and the importance of oil production for their economies is among the five highest in the sample.

In both cases the detailed quantitative examination confirmed that extraction decline episodes were entirely unanticipated. In the case of Chad, the year before the decline started production over the next five years (2005-09) was projected to increase by 25 percent relative to the pre-episode peak (2005). This optimistic projection was more that 50 percent above the actual realized production during the period.

In the case of Gabon, the year before the decline started production over the next three years (1997-99) was projected to increase by 5 percent relative to the pre-episode peak (1997). By 1999, this initial optimistic projection exceeded lower realized production by 15 percent.

Once the decline episode was underway, the degree of anticipation of the persistent decline differed considerably for Chad and Gabon. In the case of Chad, one year into the decline episode (i.e., in 2006) the extent of the subsequent 5-year fall in commodity production was largely anticipated. Specifically, the projected fall was 72 percent of the realized decline in production volumes during 2006-10. In contrast, for Gabon we find that one year into the decline episode (i.e., in 1998) production was projected to return to the level of the pre-episode peak. Thus, the fall in production was seen as fully transitory.

Overall, the in-depth exercise shows that the large and protracted production declines were entirely unanticipated a year before their start and exhibited varying degrees of anticipation after the declines were underway.

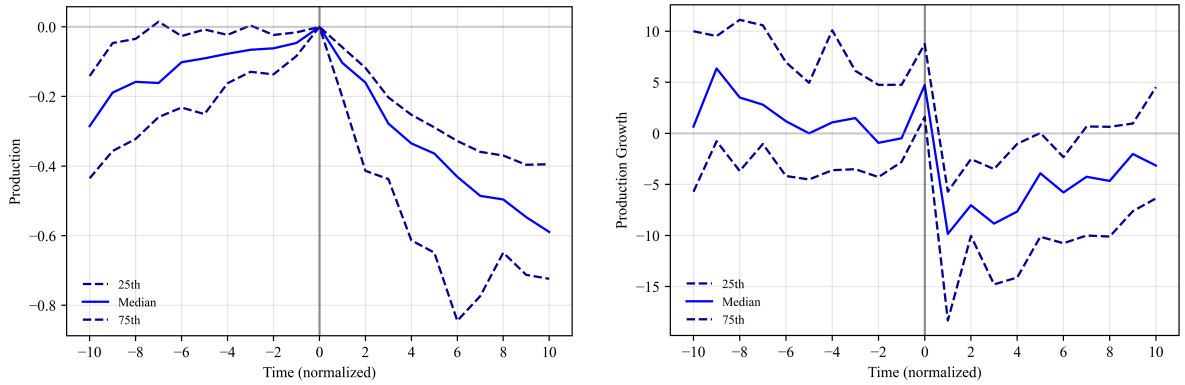
3.3 Extraction Decline Episodes: A Preview

Plotting the data of the 35 episodes gives an initial flavor of the relationship between declines in extraction and aggregate output.

3.3.1 Extractive Activity

Figure 3.2 presents the episodes of extraction declines in log-deviations from peak production before the decline and in growth rates. Time zero is the year of peak production before the start of the decline, the time window ranges from 10 years before and 15 years

after the shock.



(a) Extraction output relative to peak.

(b) Extraction output growth (in %).

Figure 3.2: Exogenous declines in fossil fuel and extractive activity.
Note: Time $t = 0$ corresponds to the peak of production and time $t = 1$ is the first year with a negative production growth rate. We normalize the log-level of production relative to the median at time zero. We present the 25th and the 75th pointwise percentiles for every horizon.

The exogenous declines have a median duration of 7 years in our sample. We measure duration from the first year of declining production after the peak at $t = 0$ to the saddle point at the permanently lower level of production. The mean duration is quite similar with 8 years. There are still some further declines in the years after up to the fourteenth year (see Figure 3.2a).

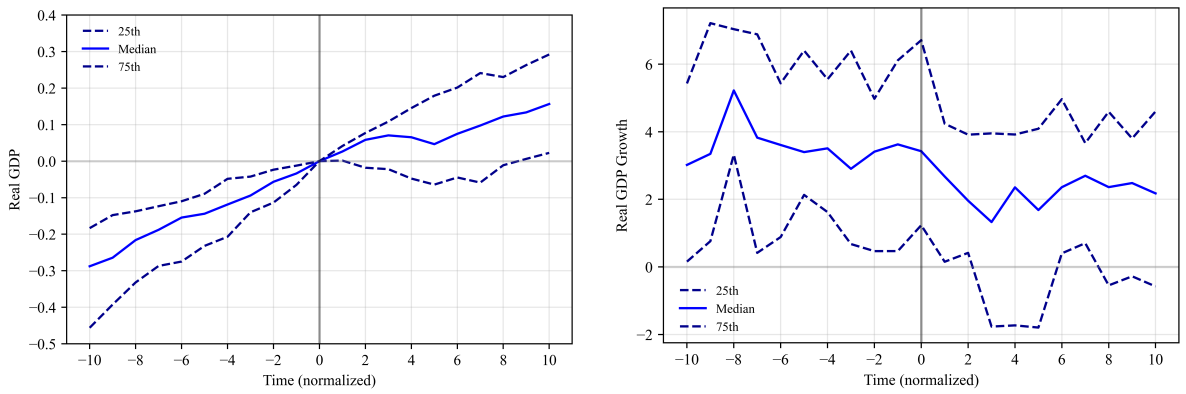
The path of the extraction declines follow a convex function with a median growth rate of -9 percent per year during the first ten years. Afterwards production declines at a median growth rate of -3 percent.[3]

The declines in our sample are skewed towards the lower end of the distribution. The gap between the 25th percentile and the median increases after the peak with a skewness factor of 2 after year zero indicating the presence of a few, very large negative production shocks. A summary of all 35 declines is provided in Section 3.10.

3.3.2 Economic Activity

Figure 3.3a shows a substantial shift in the median real GDP in response to declining extractive activity. The log-level of real GDP declines persistently, as the median does not return to its pre-shock trend even after 15 years. Accordingly, the growth rate of GDP is also persistently depressed for 15 years.

[3]As the number of observations naturally declines for a longer duration due to closed mines or missing data, the softening decline rate of production after year ten is mainly driven by a survivor bias of longer, permanent declines.



(a) Real GDP in levels.

(b) Real GDP growth (in%).

Figure 3.3: Real GDP in levels and in growth rates during the 35 episodes of extraction declines.

Notes: Time $t = 0$ corresponds to the peak of production and time $t = 1$ is the first year with a negative production growth rate. We normalize Figure 3.3a relative to the median at time zero. We present the 25th and the 75th pointwise percentiles for every horizon. The growth rate and the level do not necessarily correspond to the same country at every point in time, as the composition of the median and percentiles can differ between levels and growth rates.

3.4 Theoretical Framework

To interpret the empirical findings, we introduce a stylized two-sector model with tradable and non-tradable goods in a small open economy. In a simple endowment open economy, the response to an unexpected fall in tradable goods is linked to the shock persistence and its anticipation. According to the permanent income hypothesis, agents optimally smooth consumption in response to negative transitory shocks, decrease their consumption permanently by the full shock size in response to persistent negative shocks. As the majority of the extraction decline episodes seemed to not have been fully anticipated, this section compares the model's results for a shock that matches the typical extraction decline in the data, with and without perfect anticipation. We highlight how the lack of anticipation can delay the economic adjustment and lead to lower GDP and welfare even in the absence of other frictions.

3.4.1 Model Setup

Consider an economy populated by identical agents who maximize their lifetime utility U, defined over a sequence of an aggregate consumption basket C_t

$$U = \sum_{t=0}^{\infty} \beta^t \left[\frac{C_t^{\sigma} - 1}{\sigma} \right], \tag{3.2}$$

where $0 < \beta < 1$ is the subjective discount rate. The specified utility function exhibits a constant intertemporal elasticity of substitutions, $1/(1-\sigma)$. Consumption consists of tradable and non-tradable goods aggregated by the CES function

$$C_t = \left[(1-\gamma)(c_{N,t})^{\frac{\eta-1}{\eta}} + \gamma(c_{T,t})^{\frac{\eta-1}{\eta}}\right]^{\frac{\eta}{\eta-1}} \tag{3.3}$$

with home bias γ and η capturing the substitutability between tradable and non-tradable goods. The prices of the two goods are P_T and P_N, respectively. It is, thus, possible to define a price index

$$P_t = \left(\gamma^{\eta} P_{N,t}^{1-\eta} + (1-\gamma)^{\eta} P_{T,t}^{1-\eta}\right)^{1/(1-\eta)}, \tag{3.4}$$

household demand functions

$$C_{N,t} = \gamma \left(P_{N,t}/P_t\right)^{-\eta} C_t, \tag{3.5}$$

and

$$C_{T,t} = (1-\gamma) \left(P_{T,t}/P_t\right)^{-\eta} C_t. \tag{3.6}$$

We define all prices, P_N/P and P_T/P, in terms of the tradable good $p_t = \frac{P_{N,t}}{P_{T,t}}$.

There are two productive endowment sectors in the economy: a commodity (tradable) sector and a non-commodity (non-tradable) sector. For simplicity, non-tradable output is kept constant while tradable output is any exogenous process defined on the positive domain

$$\{y_{T,t}\}_{t=0}^{+\infty}. \tag{3.7}$$

The domestic tradable sector can be thought of as a extractive sector producing an intermediate good y_T that is exchanged in the world market at a given price for the internationally traded good c_T; given the small open economy assumption both prices can be assumed as given, which, in turn, implies that the terms of trade is exogenous (and assumed to be constant and unitary, for simplicity).

The flow budget constraint is given as follows:

$$B_t = (1+r_t)B_{t-1} + y_{T,t} + p_t y_{N,t} - c_{T,t} - p_t c_{N,t}, \tag{3.8}$$

where B_t are net foreign assets at the end of period t and p_t can be interpreted as the

real exchange rate with prices of the tradable good set as the numeraire. Market clearing implies that

$$c_{N,t} = y_{N,t}. \tag{3.9}$$

To close the model, the interest rate in the small open economy is set equal to the world interest rate, which is assumed to be in a steady state so that world interest rate equals the inverse of the discount factor:

$$r_t = r^* = \beta^{-1} - 1. \tag{3.10}$$

Finally, the economy's GDP is defined as

$$Y_t = p_t y_{N,t} + y_{T,t} \tag{3.11}$$

and the trade balance is

$$TB_t = Y_t - pc_{N,t} - c_{T,t} = y_{T,t} - c_{T,t}. \tag{3.12}$$

3.4.2 Model Results

In relation to the empirical analysis, it is instructive to compare two model scenarios. In the first scenario the empirically motivated exogenous and persistent decline in extraction output is fully anticipated as of $t = 0$. In the second scenario, the same extraction decline is implemented in the model as a sequence of transitory annual extraction output decline shocks. In a given period k, the extraction decline from $k-1$ to k is taken from data, but in all subsequent periods, i.e., from $k+1$ onward, extraction output is assumed to reverse to its $k-1$ level; i.e., $E_{t+k}[y_{T,t+k+j}] = y_{T,t+k-1}$ for $j \geq 1$.[4]

To solve the two model scenarios, we set the discount rate $\beta = 0.95$ and the intratemporal elasticity of substitution $\eta = 6$. The initial steady state net foreign asset position is set to $B_{-1} = 0$. The value of the home bias γ is calibrated such that the share of the extractive sector in total output matches the 20 percent share observed in the data in the initial steady state. Finally, for the exogenous output decline in the extraction sector

[4]While not fully consistent with rational expectations, this scenario is motivated by the lack of anticipation for extraction output declines in data (see Section 3.2.4). It can be rationalized by assuming a mining output process to be perceived as $y_{T,t} = \theta_t + \epsilon_t$ with $\theta_t = \theta_{t-1} + \eta_t$ unobservable, where ϵ and η are unobservable white noise, and the Kalman-gain evolves as $K_t = (y_{T,t-1} - \theta_{t-1|t})/(y_{T,t} - \theta_{t-1|t})$ converging to 1 in the new steady state.

we rely on empirical estimates for the 35 extraction episodes, reported in Section 3.10. The sequence of extraction declines is reported in top left chart of Figure 3.4, which summarizes the results of the two model scenarios.

In the first scenario, labelled "permanent shock", an anticipated persistent decline in mining (tradeable) output leads to an immediate fall in consumption and a depreciation of the exchange rate. The price of non-tradeable goods declines relative to the price of the internationally traded good as the country becomes poorer. This leads to a surplus in the trade. Savings increase to smooth out the consumption impact of the anticipated future decline in mining output. After an initial fall, the decline in real GDP traces the decline of the extractive sector and its trajectory is more gradual than the fall in consumption. However, consumption converges to the new steady state by falling less than GDP, as the increase in the net foreign asset position results from transitional surplus in the trade balance.

In the second scenario based on the transitory shock series, the representative consumer smoothes consumption by borrowing abroad. This induces a trade deficit but mitigates the decline in consumption. A sequence of negative unanticipated shocks, thus, leads to a partial consumption adjustment resulting in a smoother decline in consumption and output. Both eventually decline more than in the perfectly anticipated case. This is for two reasons: 1) the exchange rate depreciates only gradually and 2) the trade balance initially deteriorates. Eventually, the decrease in the net foreign asset position forces a decline in tradable consumption.

To sum up, the two model simulations reveal three key differences. First, the model predicts a trade surplus in the case of full anticipation, but a trade deficit in the unanticipated case driven by a sequence of transitory shocks. Second, in the case of a fully anticipated decline, output adjusts more gradually than consumption and vice versa for the sequence of transitory shocks. Finally, the real exchange rate depreciation is gradual and persistent under the sequence of transitory shocks.

In the next sections, we will test how our model predictions compare with the results from the empirical analysis.

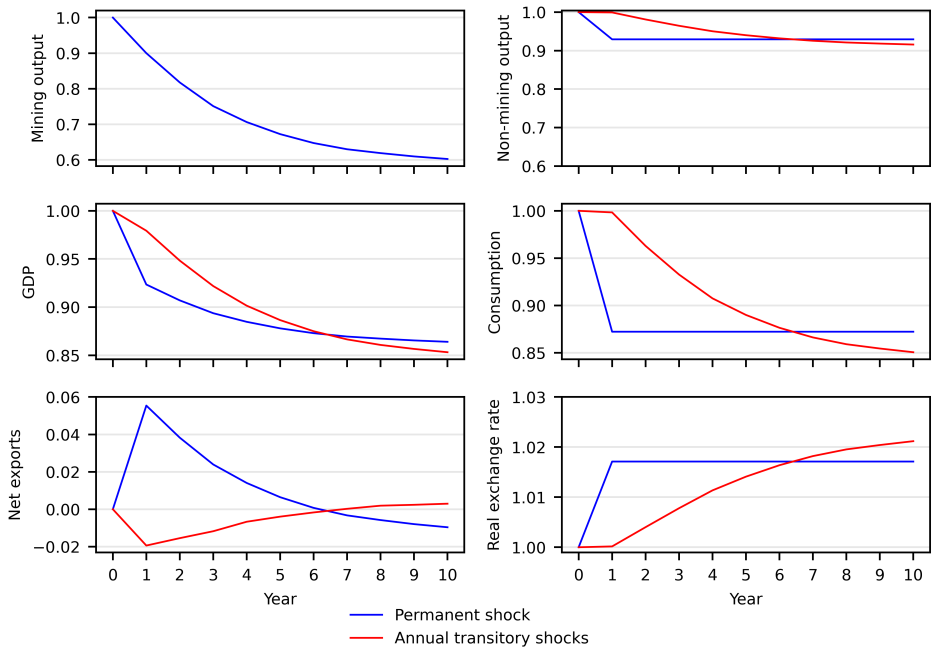


Figure 3.4: Effects of a fully anticipated permanent output decline compared to the effects of a sequence of transitory negative shocks to mining output.

3.5 Empirical Framework

For each commodity- and country-specific time series i, with $i \in \Omega_X$, a binary dummy variable, δ_{it}, takes the value of 1 if the year is part of a decline episode and 0 otherwise, where Ω_X is the set of episodes classified as exogenous in our previous section. We then instrument the change in production using the dummy in the following first-stage regression:

$$\Delta q_{it} = \alpha_i + \beta_i \delta_{it} + u_{it}, \tag{3.13}$$

where Δq_t is the percentage change in extraction for episode i at year t. While we could assess the effect of the binary dummy on the macroeconomic dependent variables directly, creating a time series based on a first stage regression allows us to calculate the effects of different sizes of production shocks. At the same time, we remove the noise related to fluctuations in production. We consider the percentage change in production as the dependent variable to circumvent any issue of non-stationarity.[5]

We use the fitted values from the first stage, $\Delta\hat{q}_{it}$, in a local projection exercise on the macroeconomic variables in line with Jorda (2005) and Stock and M. W. Watson (2018).

[5]The first stage gives the following estimated coefficients $\hat{\beta} = -s/(1-s)(\bar{\Delta q} - \bar{\Delta q}^{\star})$ and $\hat{\alpha} = \bar{\Delta q}^{\star} - \hat{\beta}$, where s is the frequency of years of production declines over the length of the sample, $\bar{\Delta q}$ and $\bar{\Delta q}^{\star}$ are the average changes in output in the overall sample and during the decline episode, respectively. The fitted values are $\bar{\Delta q}^{\star}$ during each episode.

We estimate the following equation as a baseline:

$$y_{t+h,i} - y_{t-1,i} = \alpha + \beta^h \Delta\hat{q}_{t,i} + \sum_{j=1}^{p} \Gamma_j^h y_{t-j,i} + \sum_{j=1}^{p} \Pi_j^h \Delta\hat{q}_{t-j,i} + \psi_n + \phi_t + u_{t+h,i} \quad (3.14)$$

The left-hand side represents the deviation of the log-dependent variable from its initial value. We interpret the results as cumulative percentage deviations to a shock in year t over the horizon h. The fitted values from equation 3.13 are denoted as $\Delta\hat{q}$. We include $p = 3$ lags of the dependent variable and the shock series as regressors to deal with auto-correlation following Montiel Olea and Plagborg-Møller (2021). We add country fixed effects ψ_n to account for structural differences across countries as well as time fixed effects ϕ_t to control for global price movements and other common global factors to our baseline. We cluster standard errors around every shock.

3.6 The Impact of Extraction Declines on Aggregate Output

3.6.1 The Typical Dynamics of Extraction Declines

The typical initial production shock is assumed to be a ten percent decline, which matches well the median extraction decline observed in our sample during the first year(see Figure 3.2). The initial shock has a persistent effect. The persistence of the shock is estimated, in our sample, using the following specification (see Section 3.10):

$$\Delta\hat{q}_{t+h,i} = \alpha_A + \beta_{A,h} \Delta\hat{q}_{t,i} + e_{t+h,i}; \text{ with } h = 0, \ldots, 10. \quad (3.15)$$

After the initial decline, extraction continues to decrease to an average accumulated decline of 40 percent after ten years. The typical episode is, thus, a 10 percent initial decline that accumulates to 40 percent after ten years. We will use it to assess the impact of the extraction declines on the variables of interest (see Figure 3.5) in the next sections.

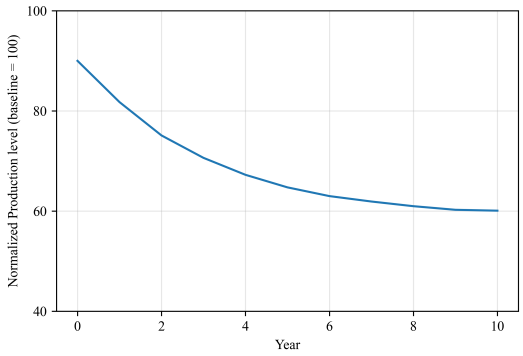


Figure 3.5: The accumulated extraction decline of an initial 10% drop in production growth accumulates to a 40% decline after 10 years.
Notes: The production decline is based on the estimated persistence of declines in our sample (as further described in Section 3.10) and matches the initial ten percent decline of the median decline episode observed in Figure 3.3b during the first year.

3.6.2 Baseline Results

During a typical extraction decline episode, real GDP falls significantly, relative to pre-shock trend, with a peak effect of seven percent after ten years (Figure 3.6). [6]

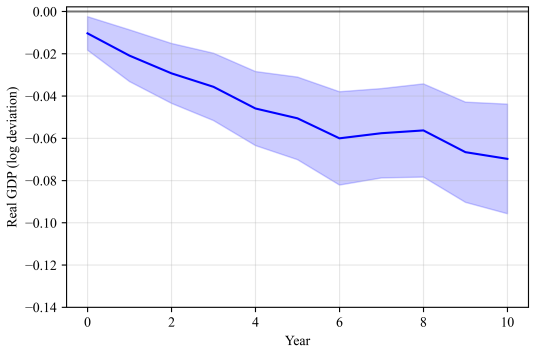


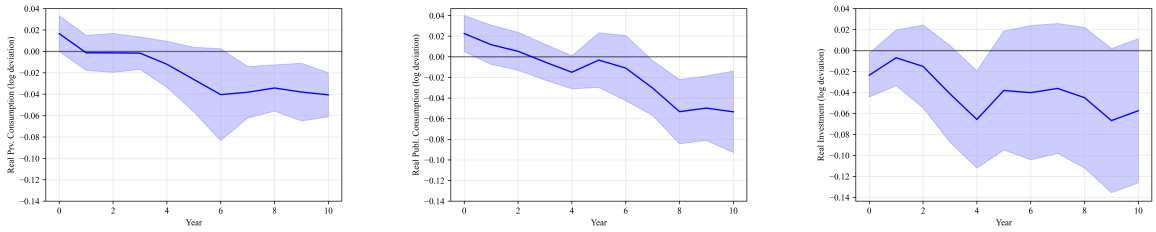
Figure 3.6: Impulse response function of real GDP to a shock episode.
Notes: The responses are generated using a shock episode normalized to an accumulated exogenous extraction decline of 40% over ten years. The point estimates are based on the beta coefficients from equation 3.14. The shaded area reports 90 percent confidence intervals.

We further analyze the real GDP response by looking at its domestic and external components, comparing results against the model. Domestically, both, real private and public consumption respond with a delay of four years to the decline and fall to levels that are -4% and -4.5% lower, respectively. The lagged responses suggest that agents misinterpreted the initial fall as an isolated one-off event and plan to smooth their consumption

[6] We use the GDP accounting identity, $Y_t = C_t + G_t + I_t + (X_t - M_t)$ to compare the relative size of sectoral drivers with the overall GDP response. We weight each sector by its respective average share of GDP in the sample. We find that the sum of point estimates, $4.1\% * 0.54 + 5.3\% * 0.16 + 5.8\% * 0.2 + (8.7\% - 3.5\%) * 0.1 = 4.8\%$, is smaller than the total GDP response of 6.9%. The difference, however, is not statistically significant. Such departure in point estimates is likely to be caused by sector share differences across countries, while the impulse response functions provide unweighted averages.

over time. Only as the decline worsens and extractive activity stays persistently at a lower level, consumption is eventually adjusted to a permanently lower level, as predicted under the perfectly anticipated case. Likewise, real investment responds with a delay. The response, however, remains insignificant over the the majority of the ten-year horizon. [7]

Overall, our results suggest that consumers and policymakers reacted in line with a series of temporary negative shocks rather than fully anticipating the entire decline in extraction output. They seem to underestimate the shock persistence and, thus, smooth aggregate effects sub-optimally.



(a) Real Prv. Consumption (b) Real Publ. Consumption (c) Real Investment

Figure 3.7: Impulse response functions of domestic sector variables to a shock episode.
Notes: Responses are generated using a shock episode normalized to an accumulated exogenous extraction decline of 40% over ten years. The point estimates are based on the beta coefficients from equation 3.14. The shaded area reports 90 percent confidence intervals.

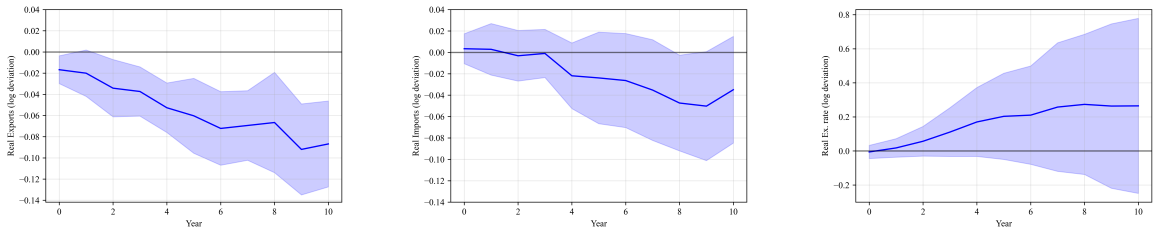
The external sector also plays an important role during the episodes. Real exports show an immediate fall, accumulating to a decline of 9% after ten years (Figure 3.8a). Real imports respond with a four-year lag and with a peak decline of 3.5% that remains insignificantly different from zero over the ten year horizon. Consequently, declining extractive activity is accompanied by a deterioration in the trade balance. Three forces are at play. First, since extraction sector output is mostly exported (directly as primary commodity or indirectly transformed into secondary commodities), the decline in mining output reduces export volumes. Second, the lagged response in consumption translates into a delayed decline in real imports. Third, possibly because of the lack of full anticipation, consumption does not decline enough, supporting demand for imports and inducing an increase in borrowing, which in turn allows the trade balance to worsen. This helps to mitigate the decline in consumption.

The real exchange rates should play the role of a shock absorber. It depreciates per-

[7] We also assess employment responses as shown in section 3.10. Employment falls slightly in response to the shock and is only statistically significant between the six and eighth year horizon. The muted response could be driven by the high capital intensity of the sector. At the same time, employment data for low and middle income countries is not reliable, which may introduce an attenuation bias.

sistently by approximately 20 percent after four years and remains at its lower level thereafter, even though the response is not precisely estimated (see Figure 3.8c). The depreciation indicates that the economy becomes poorer, as the price of the non-tradable output declines relative to the price of the fossil fuel. At the same time, the decline could potentially limit the fall in exports by stimulating non-mining exports and re-allocating resources towards other tradable sectors. Our results suggest, however, that the induced reallocation channel is too weak to reverse the negative consequences of the mining decline. Moreover, countries under fixed exchange rate regimes will face strong devaluation pressures in response to large and persistent declines in extractive sectors.

Overall, the comparison of empirical and theoretical results suggests that governments and firms probably misjudged the impact and persistence of declines in extractive activity. The delay in responses indicates that agents either perceived shocks to be transitory or learned slowly as the decline in the extraction sector unfolded.



(a) Real Exports (b) Real Imports (c) Real Ex. Rate

Figure 3.8: Impulse response functions of external sector variables to a shock episode.
Notes: Responses are generated using a shock episode normalized to an accumulated exogenous extraction decline of 40% over ten years. The point estimates are based on the beta coefficients from equation 3.14. The shaded area reports 90 percent confidence intervals.

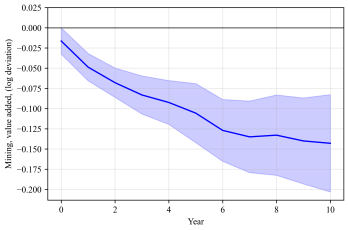
3.6.3 Effects Across Industries

We decompose the real GDP response by industry. Figure 3.9 displays the response for the three largest sectors in our sample, namely mining, manufacturing and services. The mining sector drops by 2% at the time of the shock, $t = 0$, and falls by 14.5% after ten years. The convex decline mimics the fall in production shown in Figure 3.2a. Moreover, the immediate drop in mining at time $t = 0$ provides an explanation for the corresponding drop of real GDP on impact.

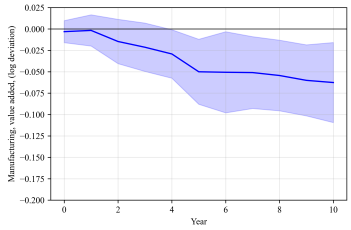
The manufacturing and service sectors react with a lag of three years each. Their value added falls by 6% and 5.5%, respectively. As domestic aggregate consumption and investment react with a similar lag, the lag in the response of manufacturing and service

is in line with a domestic demand-driven channel.

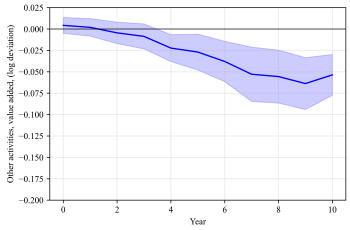
An additional explanation is that manufacturing and services are likely related to the extraction sector both in terms of inputs of machinery and business services, respectively. Both sectors react with a lag as mining output gradually falls over several years in most episodes. The decline in extractive activity, in fact, may have spillover effects on these industries, reflecting its large footprint on the economy that goes beyond its simple output share.



(a) Mining,value added



(b) Manufacturing, value added



(c) Services & others, value added

Figure 3.9: Impulse response functions of real value added (in constant 2015 local currency units) by sector to a shock episode.
Notes: Responses are generated using a shock episode normalized to an accumulated exogenous extraction decline of 40 percent over ten years. The point estimates are based on the beta coefficients from equation 3.14. The shaded area reports 90 percent confidence intervals. The IRFs of the remaining sectors are shown in Section 3.10.

3.6.4 Heterogeneity Across Countries

We test for potential heterogeneity in responses across countries by augmenting the local projection with multiple interaction terms for per capita income and extractive sector size. We also control for the share of the manufacturing sector in total output and institutional quality.

Income and Extractive Sector Size Effects:

As a first exercise we control for the heterogeneity in the shock size and interact the decline in quantities with the share of the value added of the extractive sector in total value added MS_i averaged over the 10 years preceding each episode:

$$y_{t+h,i} - y_{t-1,i} = \alpha + \beta^h \Delta\hat{q}_{t,i} + \gamma^h(\Delta\hat{q}_{t,i} * MS_i) + \theta^h MS_i + \sum_{j=1}^{p} \Gamma_j^h X_{t-j,i} + \psi_n + \phi_t + u_{t+h,i} \tag{3.16}$$

Figure 3.10 shows the response for small and large extractive sectors corresponding to the 10th and 90th percentile in our sample—that is, a mining share of 2.5% and 58%, respectively.[8] Countries in our sample with a relatively small extractive sector experience relatively muted declines of 3.7% in real GDP after ten years. For countries with large extractive sectors the decline is significantly larger with an average 15.3% fall in real GDP after ten years. The impact of the share of the extractive sector, however, is less than proportional, since doubling the mining share increases the impact on real GDP on average by only about 40 percent.[9]

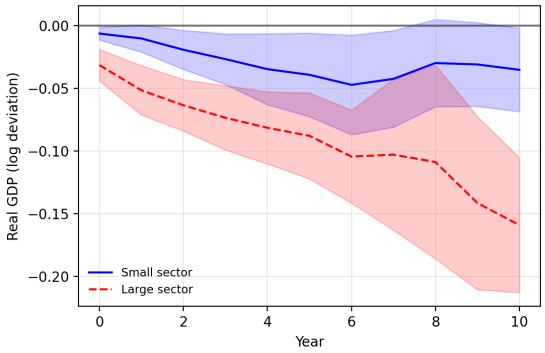


Figure 3.10: Impulse response function of real GDP interacted with the size of the resource sector to a shock episode.
Notes: Responses are generated using a shock episode normalized to an accumulated exogenous extraction decline of 40% over ten years. The point estimates are based on the coefficients from equation 3.16. The shaded area reports 90 percent confidence intervals.

We assess the impact countries' different income levels on GDP responses. We include an interaction term between the shock series and a binary dummy I_n that indicates the country's per capita income level. The binary dummy variable takes the value of 0 for high income and 1 for middle and low income countries based on the country classification by the World Bank (2022). The interaction term allows us to assess the additional shock effect for different income levels.[10] We run the following augmented local projection:

$$y_{t+h,i} - y_{t-1,i} = \alpha + \beta^h \Delta\hat{q}_{t,i} + \gamma^h(\Delta\hat{q}_{t,i} * I_n) + \sigma^h(\Delta\hat{q}_{t,i} * MS_i) + \rho^h(\Delta\hat{q}_{t,i} * I_n * MS_i) + \theta^h MS_i + \sum_{j=1}^{p} \Gamma_j^h X_{t-j,i} + \psi_n + \phi_t + u_{t+h,i} \quad (3.17)$$

[8] To give a sense of proportion, mining accounted for 1.2 percent of value added in 2019 in the U.S., while for a primary commodity importer such as Italy mining and quarrying account for about 0.2 percent of value added.

[9] It is possible that a bigger extractive sector might also be more diversified across commodities.

[10] Note that since I_n is constant over time, we do not include it as a separate regressor as it is captured by the country fixed effects. We further move all lagged controls into X.

Figure 3.11 shows that the declines in extractive activity are particularly detrimental for low-income countries with large extractive sectors.[11] While the income level plays a relatively minor role for countries with small extractive sectors, the impact on GDP is especially pronounced for low-income countries with large extractive sectors.[12] In response to a decline in extraction of 40% after ten years in a sector that makes up about 58% of aggregate value added, real GDP in low-income economies falls by 30%. This compares to a 10% decline in high-income countries.[13] The initial impact is stronger in high-income countries (with a large extractive sector). Three years after the shock, instead, the ranking flips and the difference becomes pronounced over time, suggesting that low income countries are not able to cushion the detrimental effect on aggregate output as much as high income countries. This may indicate the lack of means to stabilize or transform the economy away from resource-linked activities or could point the role of institutions (see Section 3.7).

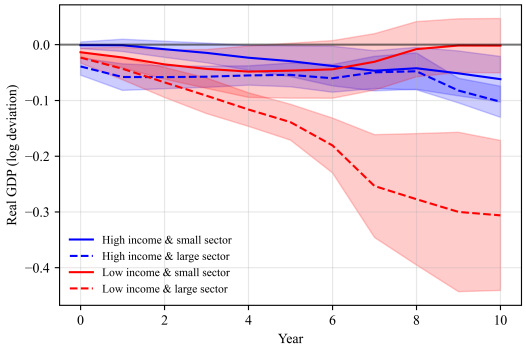


Figure 3.11: Impulse response function of real GDP interacted with per capita income and sector size to a shock episode.
Notes: Responses are generated using a shock episode normalized to an accumulated exogenous extraction decline of 40% over ten years. The point estimates are based on the coefficients from equation 3.16. The shaded area reports 90 percent confidence intervals.

Manufacturing Sector Size Effects:

We estimate the effects of the shock episodes on real GDP controlling for both the mining and the manufacturing share of value added. As the extractive sector declines and the exchange rate depreciates, resource reallocation towards non-mining export sector may be able to buffer the shock impact. When controlling for manufacturing sector sizes, data availability becomes a limiting factor. In particular, the value added of

[11] Note that the small sector size corresponds to a 10th percentile in our sample (5 percent for HICs and 2.5% for LICs) and the large sector size is defined by the 90th percentile in our sample, which equals 58% for HICs and 50% for LICs.

[12] Section 3.10 displays the response under a further linearity restriction implying a linear effect of the sector size on real GDP.

[13] There are 11 high and 24 low-income countries in the sample (see section 3.10).

manufacturing sectors is available for only 25 shock episodes. Given the small sample size the statistical power falls substantially, especially when controlling for multiple interaction terms. In order to maintain a high level of statistical power, we control for mining and manufacturing with a mining-manufacturing sector ratio $MR_i = \frac{MS_i}{Manf_i}$ in the following regression:

$$y_{t+h,i} - y_{t-1,i} = \alpha + \beta^h \Delta\hat{q}_{t,i} + \gamma^h(\Delta\hat{q}_{t,i} * MR_i) + \theta^h MR_i + \sum_{j=1}^{p} \Gamma_j^h X_{t-j,i} + \psi_n + \phi_t + u_{t+h,i} \tag{3.18}$$

Figure 3.12 shows the real GDP response for countries with small and large extractive sectors relative to the manufacturing sector. The results indicate that a larger manufacturing sector can buffer the shock to extraction. The small mining-manufacturing ratio corresponds to the 10th percentile in our sample and represents a extractive sector that is 30% as large as the manufacturing sector. The large mining-manufacturing ratio corresponds to the 90th percentile in our sample and represents a extractive sector that is eight times as large as the manufacturing sector.

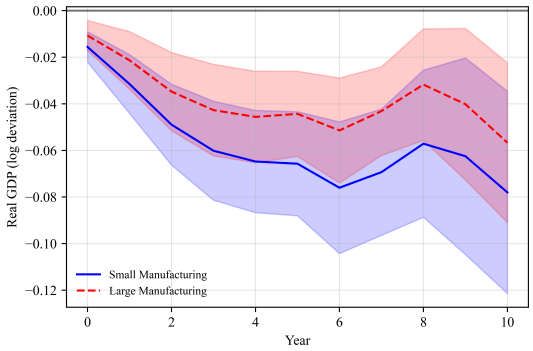


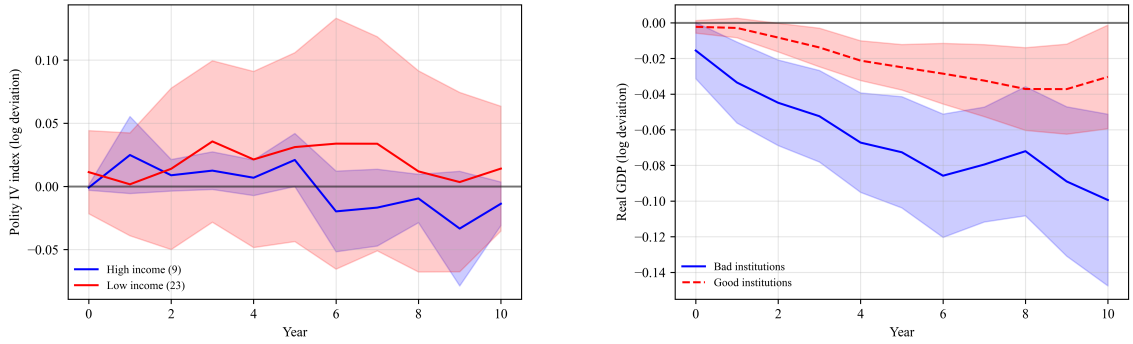
Figure 3.12: Impulse response function of real GDP interacted with the extractive sector and the manufacturing sector share of value added to a shock episode.
Notes: Responses are generated using a shock episode normalized to an accumulated exogenous extraction decline of 40% over ten years. The shaded area reports 90 percent confidence intervals.

3.7 The Role of Institutions

An extensive literature highlights that resource booms lead to a deterioration in the quality of institutions, negatively affecting other sectors of the economy. In the reverse case, does a decline in extraction lead to better institutions or is there a legacy effect from rent seeking? To test this hypothesis, we follow Acemoglu et al. (2019) and use a data-

set on institutional quality from the Center for Systemic Peace (2022) as a dependent variable in Equation (3.14).

We find that persistent declines in extraction do not cause improvements in institutional quality in neither high nor low-income countries (see Figure 3.13a). The impact is statistically insignificant from zero. This means that there is a potential asymmetry, if institutions were damaged during the resource boom, a decline in extraction is not enough to improve them. Once institutions are damaged they are hard to fix.



(a) Institutional quality

(b) Real GDP and institutional quality

Figure 3.13: Impulse response function with the institutional quality indicator (-10 to 10) to a shock episode.
Note: Figure 3.13a shows the impulse response of institutional quality to a shock episode accounting for the income per capita level. Figure 3.13b shows the impulse response of real GDP to a shock episode accounting fr the institutional quality level. Responses are generated using a shock episode normalized to an accumulated exogenous extraction decline of 40% over ten years. The shaded area reports 90 percent confidence intervals.

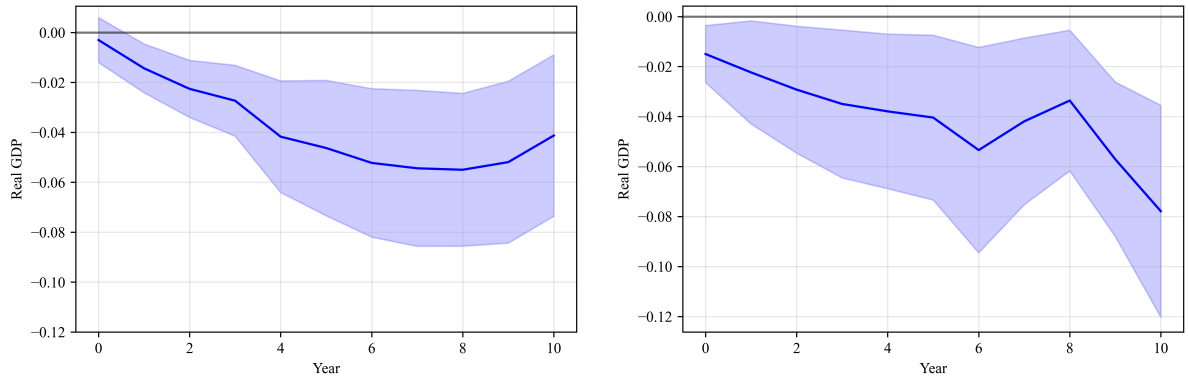
We also investigate whether the quality of institutions before the extraction decline can affect the economic impact of such a decline. This could explain why high income countries, having on average better institutions, fare better during extraction declines. We include the variable for institutional quality as a separate interaction term in Equation (3.14). Figure 3.13b shows the effects of the extraction decline episodes accounting for different levels of institutional quality. Five years after the shock, the difference in real GDP between countries with high and low institutional quality is about 5 percentage points. While a democratic country with high institutional quality (i.e. an indicator of 10) experiences a decline in real GDP of 2.5%, an autocratic country with bad institutions (i.e. an indicator of -10) exhibits a decline of 10% after ten years. Finding an explanation for this is beyond the scope of the current paper. It is possible that higher institutional quality leads to a more diversified economy or to more promptly enact structural policies to address the shortfall in mining output and revenues, including

fiscal revenues. Our results suggest that institutional quality plays a role in determining the impact of the declining extractive activity on the economy.

3.8 Robustness

3.8.1 Sub-Sample Analysis

Our results are robust to sub-samples of shock episodes: one that includes all non-oil shock episodes, and one with the oil shock episodes only. Figure 3.14 shows the impulse response of real GDP to the shock episodes in the two sub-samples, respectively. In both cases, the shock consisting of a 40% decline in extraction after ten years causes a real GDP slow down of between five and six percent after six years. While the impact levels off for all commodities ex. oil, the negative impact becomes even more pronounced based on the oil shock episodes only after 10 years.



(a) Real GDP (excl. oil and gas shocks)

(b) Real GDP (only oil and gas shocks)

Figure 3.14: Impulse response functions of sub-sample Real GDP to a shock episode.
Notes: Figure 3.14a is based on 15 decline episodes in metals and Figure 3.14b is based on 20 decline episodes in oil and natural gas. Responses are generated using a shock episode normalized to an accumulated exogenous extraction decline of 40% over ten years. The shaded area reports 90 percent confidence intervals.

We also check the robustness of our results for excluding one of the 35 declines episodes at a time. This allows us to investigate whether the results are driven by particular decline episodes. The results in presented in Section 3.10 show that they are robust to this sensitivity check.

3.8.2 Anticipation Effects

To assess the potential differences between anticipated and unanticipated extraction decline episodes, we run the local projection exercise including a binary dummy for anticipation. Figure 3.15 shows that anticipated declines lead to an insignificantly smaller response in point estimates over the 10-year horizon than for the unanticipated ones. This suggests that even the anticipated episodes are hard to predict. This leaves little time for the economy to adjust before the shock is realized. This seems a substantial difference with the initial stage of the resource curse where, after large mineral deposits are discovered, the increase in resource extraction is much more predictable as in Arezki et al. (2017).

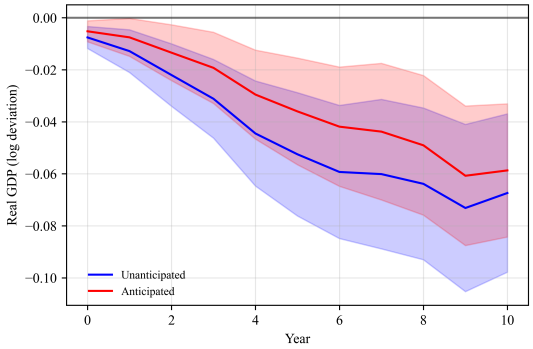


Figure 3.15: Impulse response functions of GDP interacted with a dummy for anticipation to a shock episode.
Notes: Responses are generated using a shock episode normalized to an accumulated exogenous extraction decline of 40% over ten years. The shaded area reports 90 percent confidence intervals. We include 27 shock episodes in this local projection, 24 are unanticipated and four are anticipated. There is no data available for the remaining seven.

3.8.3 Long-Term Results

Impulse response functions for a longer time horizon show that the impact on real GDP is quite persistent. After twenty years, real GDP is still 4% lower due to the decline in extraction. The persistence indicates that there is no reversed resource curve even 20 years after the initial shock.

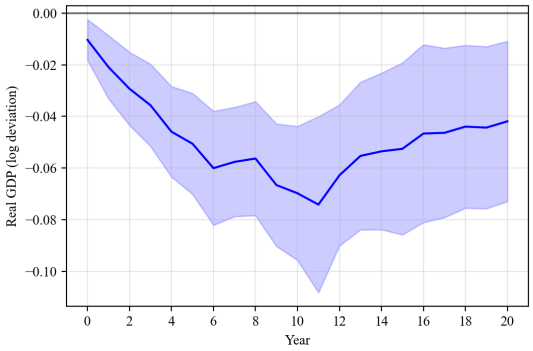


Figure 3.16: Average impulse response of real GDP over 20 years to a shock episode.
Notes: Responses are generated using a shock episode normalized to an accumulated exogenous extraction decline of 40% over ten years. The point estimates are based on the coefficients from equation 3 with a horizon of 20 years. The shaded area reports 90 percent confidence intervals.

3.8.4 Synthetic Counterfactual Method

We employ the synthetic counterfactual method (SCM) by Abadie, Diamond, and Hainmueller (2010). While local projections provide us with an average prediction for shock responses via IRFs, synthetic counterfactuals works complementary by allowing us to summarize the shock response on an country-average and case-by-case basis.

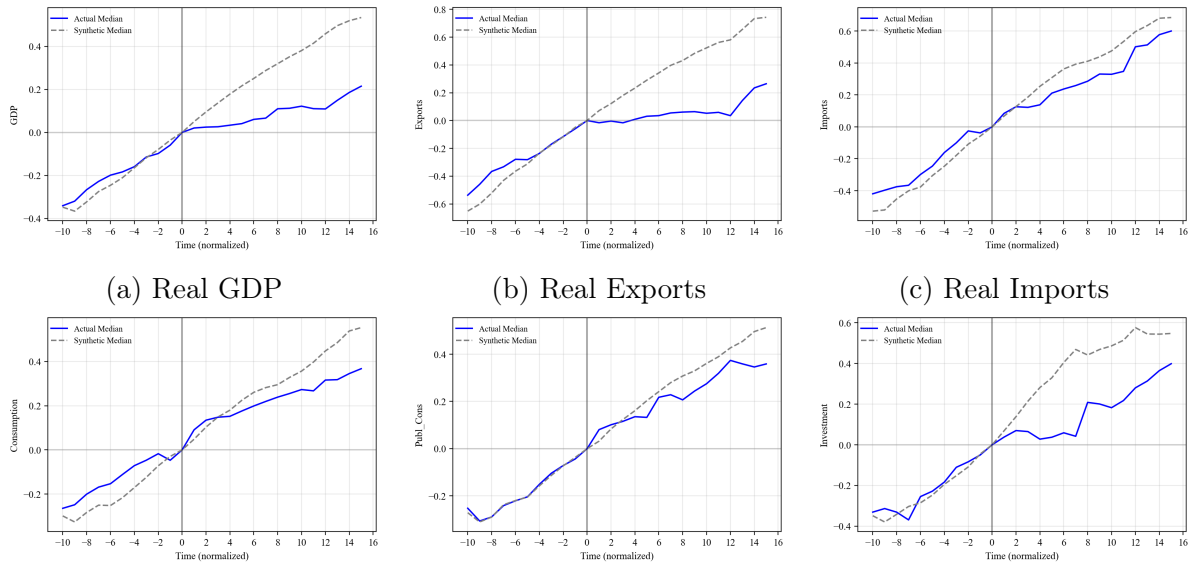
The synthetic control method estimates the impact of an intervention by comparing the actual evolution of a variable for a country to the development of the same, aggregate variable in a synthetic control group.[14] This is particularly helpful when a suitable single control does not exist. Instead, the method employs a combination of all countries worldwide that are not affected by similar shocks at that time and that are similar to the treated economy along some dimensions (e.g. also a producer of the same commodity or similar levels of GDP). The synthetic counterfactual drawn from this pool for each treated economy represents a weighted average counterfactual of control countries that together mirror the development of the recipient in the pre-shock episode.

To summarize the results across the whole sample, we use the difference between the synthetic and actual development to measure the median effect over time. To ensure that this gap is based on a well-fitted counterfactual, we trim the observations to only include countries with a root mean square prediction error (RMSPE) below the 75th percentile. Section 3.10 showcases the country-specific SCM results based on one shock episode.

The results summarized in Figure 3.17 confirm our previous findings. Median real GDP

[14] See Section 3.10 for more details on the estimation equations.

decreases by approximately 15 percent after six years relative to the synthetic counterfactual.[15] The decline in real GDP is matched by the difference in declines between real exports and imports over time. Employment deviates from its synthetic counterfactual with a lag of six year before it returns to its counterpart in year 14. Real consumption falls, however, less so then real GDP. Interestingly, investment declines just as much as GDP. However, it is important to mention that the overall fit of investment in the SCM may reduce the reliance on its estimates.



(a) Real GDP (b) Real Exports (c) Real Imports

(d) Real Prv. Consumption (e) Real Publc. Consumption (f) Real Investment

Figure 3.17: Actual versus synthetic counterfactual medians over time.

3.9 Conclusion

This paper is the first to empirically examine the macroeconomic effects of persistent contractions of fossil fuel and other mine production. Based on a new data-set on the extraction of 13 different minerals for 122 countries worldwide starting in 1950, we identify 35 episodes of large and persistent declines in extractive activity that are exogenous to a country's macroeconomic activity.

Using local projects methods, we find that declines in extractive activity lead to a significant and persistent negative effect on real GDP. The real exchange rate depreciates slowly and not enough to offset the decline in exports. This leads to a deterioration of the trade balance. Private and public consumption fall but with a lag. There are, in fact, significant, negative spillover effects of declining extractive activity on the manufacturing

[15] In comparison, GDP in Figure 3.3a is about 13 percent lower than its linear trend after six years.

(and service) sectors. Results are qualitatively consistent with a classical small-open-economy model, in which agents learn only slowly about the persistence of the shock over time.

Results show significantly larger impacts for low-income countries when controlling for the sector size. This suggests a greater ability of high-income countries to buffer the detrimental effects on economic activity. A possible explanation is that high income countries exhibit on average better institutions, which could mitigate the impact of the extraction decline. Declines in extraction seem to not improve the quality of institutions, possibly perpetuating the negative impact of the resource sector even after the resource sector looses importance.

Our findings suggest that enhancing climate policy certainty at the global level would enable a smoother transition for countries that heavily rely on fossil fuel exports. Such certainty could provide incentives to initiate an early and gradual adjustment to the declining extraction of fossil fuels. Therefore, it is vital for policymakers to provide clear and consistent signals so that fossil fuel production declines are well anticipated. In doing so, the cumulative utility of citizens as well as the national GDP are maximized as compared to a scenario where the energy transition is not fully anticipated, nor internalized.

Our paper and the new data-set suggest several potential areas for future research. For example, one could empirically study to what extent specific institutions and policies mitigated some of the macroeconomic effects of the extraction declines in the historical episodes. In addition, a broader data-set could examine the impact of the decline in extractive activity on the fiscal accounts. Such an analysis could shed light on the extent to which declining extractive activity affects government revenues, public spending priorities, and overall fiscal sustainability. Lastly, one could also study the distributional impacts of the green energy transition to better understand the feasibility and sustainability of public policies aimed at embracing it. If the losers of such a transition are properly compensated from the net benefit of better policy, then a more stable and net beneficial state could potentially be reached as minimal opposition will occur.

3.10 Appendix - Empirical Analysis

Summary Statistics: Entire Data-set

Commodity	# of Decline Episodes	Decline of Extraction Quant. (% of Peak)	Decline of Extraction Value (Mil. USD)	Duration (Years)	Annualized Decline Value (% of GDP)
Bauxite	10	-69	-47	6	-3.4
Brown Coal	2	-76	-6	7	0.0
Copper	17	-80	-226	9	-3.8
Crude Oil	21	-47	-8290	6	-19.4
Gold	9	-50	-379	5	-1.7
Hard Coal	11	-77	-113	5	-0.2
Lead	30	-75	-25	9	-0.1
Natural Gas	7	-60	-6080	7	-0.5
Nickel	3	-46	-49	3	-0.5
Iron	14	-79	-55	5	-1.5
Silver	10	-64	-24	9	-0.2
Tin	11	-85	-22	11	-0.4
Zinc	9	-60	-153	6	-0.2
Total	154				

Table 3.2: Summary of episodes of declining extractive activity by commodity (in averages).
Note: Columns 2 and 3 show the average total decline over all the years of each episode. Column 4 shows the average duration of the episodes. Column 5 (the average ADV) is the result of dividing column 3 by column 4 and the GDP at time $t = 0$ for every shock.

Summary Statistics: 35 Extraction Decline Episodes

1	2	3	4	5	6	7
Country	**Commodity**	**Year**	**Duration (Years)**	**ADV (% of GDP)**	**Reason**	**Anticipated**
Kuwait	Oil	1979-1982	4	23.3	Policy change	No
Brunei Darussalam	Oil	1979-1981	3	13.5	Policy change	NA
Chad	Oil	2005-2014	10	4.1	Depletion	No
Suriname	Bauxite	1974-1977	3	3.8	Policy change	NA
Gabon	Oil	1997-2002	6	3	Depletion	No
Brunei Darussalam	Oil	2006-2018	13	2.35	Depletion	No
Eswatini	Iron	1975-1979	5	2.2	Depletion	NA
Equatorial Guinea	Oil	2012-2019	8	2.2	Depletion	No
DR Congo	Zinc & Copper	1988-1994	7	1.9	Depletion	No
Guyana	Bauxite	1975-1982	8	1.7	Policy change	NA
Norway	Oil	2003-2013	11	1.5	Depletion	Yes
Zambia	Copper	1988-1994	7	1.5	Policy change	No
Jamaica	Bauxite	1981-1986	5	1.3	Policy change	No
South Africa	Silver, Lead & Gold	1994-2001	8	1	Depletion	No
Trinidad and Tobago	Oil	2006-2019	14	1	Depletion	No
Bolivia	Tin	1981-1986	6	0.7	Depletion	No
Syria	Oil	2003-2010	8	0.7	Depletion	No
Indonesia	Oil	2000-2007	8	0.5	Depletion	No
Fiji	Gold	2004-2007	4	0.5	Depletion	No
Mexico	Oil	2004-2009	6	0.5	Depletion	No
Peru	Oil	1985-1991	7	0.4	Depletion	No
Zimbabwe	Nickel	1998-2004	7	0.4	Policy change	No
Tunisia	Oil	2007-2017	11	0.3	Policy change	No
Denmark	Gas	2008-2013	6	0.3	Depletion	Yes
South Africa	Gold	1970-1975	6	0.3	Depletion	No
Denmark	Oil	2004-2019	16	0.2	Depletion	No
United Kingdom	Oil	1986-1990	4	0.2	Depletion	Yes
Myanmar	Lead	1973-1976	4	0.2	Depletion	NA
Tunisia	Oil	1992-2001	10	0.2	Depletion	Yes
Netherlands	Gas	2013-2019	6	0.2	Depletion	NA
United Kingdom	Oil	1999-2015	16	0.2	Depletion	No
Nicaragua	Copper	1968-1978	11	0.2	Depletion	NA
Zambia	Lead	1988-1994	7	0.1	Depletion	No
Australia	Oil	2000-2006	7	0.1	Depletion	No
Philippines	Copper	1981-2002	21	0.1	Depletion	No

Table 3.3: Summary of the 35 episodes of exogenous extraction declines.

Notes: Countries sorted by the estimated value of the extraction decline as share of GDP (ADV). Anticipation indicates whether the initial decline was anticipated in IMF article IV consultation reports. There are no reports for 6 out. In the case of the Netherlands, natural gas production was not discussed.

Model: First-Order Conditions

This appendix provides the first-order conditions for the two-sector model as well as its calibration. Equation (3.19) shows the intra-temporal Euler equation relating the choice between non-tradable and tradable consumption to their relative price p with home bias γ as well as the substitutability between non-tradable and tradable goods η. Equation (3.20) shows the inter-temporal Euler equation relating current consumption to future consumption via their relative price r and the discount factor β.

$$\left(\frac{c_{T,t}}{c_{N,t}}\right)^{\frac{1}{\eta}} = p * \frac{\gamma}{1-\gamma} \tag{3.19}$$

$$\left(\frac{c_{T,t+1}}{c_{T,t}}\right)^{\frac{1}{\eta}} = \beta * (1+r) \tag{3.20}$$

Synthetic Control Methodology

Following Abadie and Gardeazabal (2003), let y_{it} denote the observed variable of interest with the set of determinants Z_{it}, where i denotes the shock ($i = 1$ corresponding to the treated country and ($i = 2, \cdots, N$ to the control countries) at time $t \in [-10, \cdots, T]$. The synthetic counterfactual is a weighted average of y_{it} (with $i = 2, \cdots, N-1$ and $t < 0$) computed to match y_{1t} accounting for the determinants Z. The set of weights is $W = (w_2, \cdots, w_{N+1})$ with $w_i \geq 0$ for $(i = 2, \cdots, N-1)$ and $\sum_{i=2}^{N-1} w_i = 1$. Assuming no anticipation effects of production shocks, the synthetic counterfactual $\sum_{i=2}^{N-1} w_i y_{it}$ equals the treated country y_{1t} during the pre-shock episode as follows

$$\sum_{i=2}^{N-1} w_i y_{it} = y_{1t},$$

where

$$\sum_{i=2}^{N-1} w_i Z_{it} = Z_{1t}.$$

The optimal vector of weights W^* solves

$$\min (S_1 - S_C W)' M (S_1 - S_C W),$$

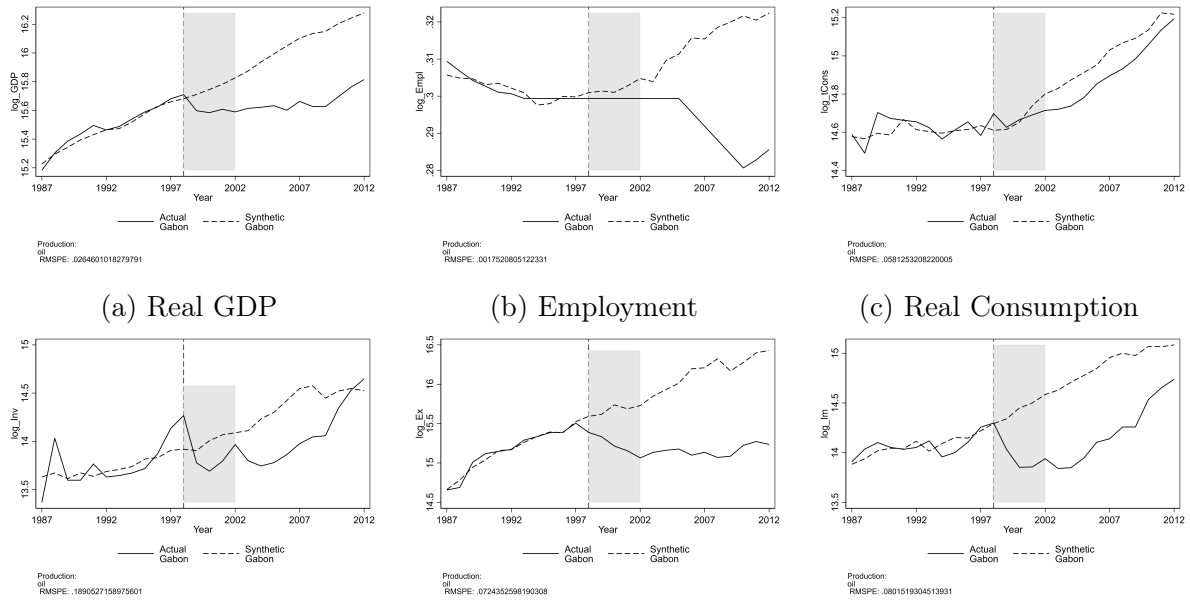
$$\text{s.t. } w_i \geq 0$$

$$\text{and} \quad \sum_{i=2}^{N-1} w_i = 1,$$

where S_1 is the vector $(k \times 1)$ of the treated country characteristics (k) in the pre-shock episode, S_C is the vector $(k \times N)$ of the same variables for the control nations and M is a $(k \times k)$ symmetric, semi definite matrix measuring the relevance of the variables included in S_1 and S_C. W^* minimizes the squared error of the pre-shock distance between the vector of the treated country variables and the vector of the synthetic control variables.

Synthetic Control Method Example - Country Results

We use Gabon as an example for a by-country analysis. The synthetic counterfactual for GDP is based on a weighted average of eight countries that are also oil producers and together mimic both, Gabon's pre-shock oil production as well as real GDP.[16] Section 3.10 shows that real GDP drops by 16 percent after two years compared to the counterfactual. The gap between actual and synthetic GDP widens to a peak size of 52 percent after eleven years.



(a) Real GDP (b) Employment (c) Real Consumption

(d) Real Investment (e) Real Exports (f) Real Imports

Figure 3.18: Synthetic counterfactual for Gabon.
Note that the shaded area shows the duration of the oil production shock.

*

Estimated Shock Persistence

[16]The weighted countries are: Angola, 8.1%; Ecuador, 18.8%; India, 29.4%; Iran, 5%; Italy, 18.9%; Saudi Arabia, 0.9%; Thailand, 15.1% and Vietnam, 3.7%.

Following Francesca Caselli et al. (2022), we estimate the persistence of shocks in our sample based on the following specification:

$$\Delta\hat{q}_{t+h,i} = \alpha_A + \beta_{A,h}\Delta\hat{q}_{t,i} + e_{t+h,i}, \tag{3.21}$$

with the fitted shock series of extraction declines $\Delta\hat{q}$ at time t, horizon h, and commodity-episode i, the coefficient for the constant α_A, the autocorrelation coefficient β_A^h and the regression residual e. Figure 3.19 illustrates the corresponding shock persistence underlying the local projection based on the coefficient β_A^h. The y-axis reports the unit of the auto-correlation coefficient. A shock at time t tends to persist throughout subsequent years with an average duration of eight years.

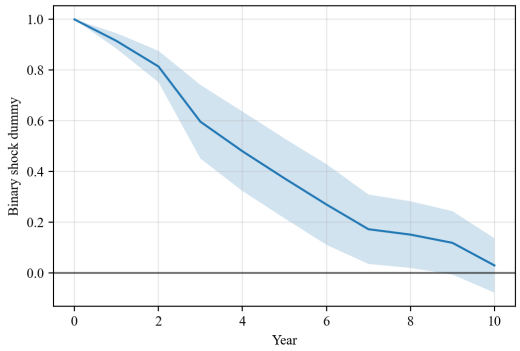


Figure 3.19: Estimated persistence (auto-correlation) of the binary shock dummy.

We can iteratively define the baseline production level, $Y_t \equiv Y_{t-1}\Delta\hat{q}_{t+h,i}$, to accumulate the shock persistence to display the decline of the (trend) production level over time by multiplying the β_A^h coefficient iteratively with :

$$Y_{t+h} = Y_{t+h-1} * (\beta_{A,h} + \alpha_A) = Y_{t-1} * (\beta_{A,h} + \alpha_A) * \cdots * (\beta_{A,0} + \alpha_A) \tag{3.22}$$

where Y_{t+h} is the extraction level over the horizon h, β_A^h is the auto-correlation coefficient of the shock from Equation (3.21).

Employment Response

This chart shows the average impulse response of the employment rate to the shock episodes. The employment rate is defined as the number of employed people in the labour force (i.e. older than 15 and younger than 65) over the population. The data is not reliable for many of the middle and low-income countries in our sample.

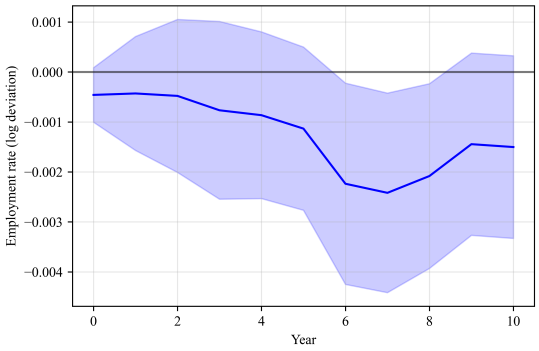


Figure 3.20: Impulse response functions of the employment rate to a shock episode.
Notes: Responses are generated using a shock episode normalized to an accumulated exogenous extraction decline of 40% over ten years. The shaded area reports 90 percent confidence intervals.

Linearity Restriction

This chart displays the impulse responses for real GDP when restricting the effect of the sector size as shown in the following equation:

$$y_{t+h,i} - y_{t-1,i} = \alpha + \beta^h \Delta(\Delta \hat{q}_{t,i} * MS_i) + \sum_{j=1}^{p} \Gamma_j^h X_{t-j,i} + \psi_n + \phi_t + u_{t+h,i}$$

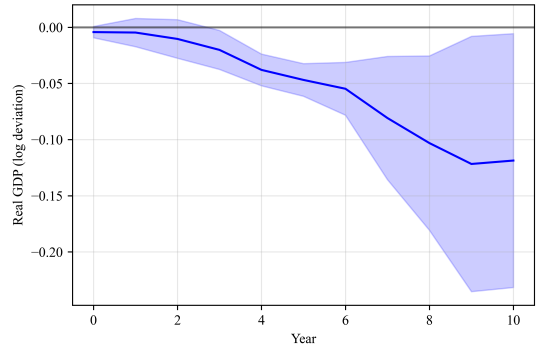


Figure 3.21: Impulse response function of real GDP to a shock episode with linearity restriction.
Notes: Responses are generated using a shock episode normalized to an accumulated exogenous extraction decline of 40% over ten years. The shaded area reports 90 percent confidence intervals.

Sub-Samples

This chart shows the impulse responses for real GDP excluding one shock episode at a time.

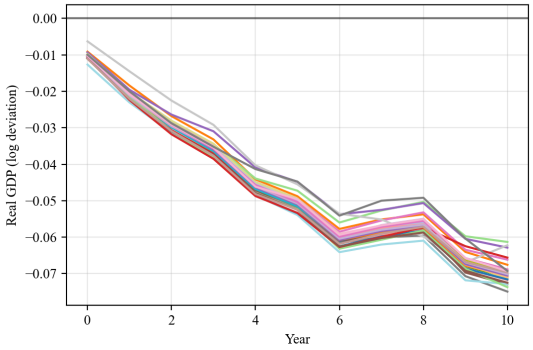


Figure 3.22: Impulse response functions excluding one decline episode a time.
Notes: Responses are generated using a shock episode normalized to an accumulated exogenous extraction decline of 40% over ten years. The shaded area reports 90 percent confidence intervals.

Methodology to Examine Anticipation

The purpose of studying anticipation or lack thereof by tracking actual commodities production data, as well as forecasts included in IMF Article IV staff reports and updates to those data and forecasts, is to see whether commodity supply shocks were internalized.

We identify 35 episodes of persistent shocks to commodities extraction. These shocks vary in geography, duration, and affected commodities. For each shock, we examine IMF Article IV staff reports to determine whether the macroeconomic forecasts prepared by the IMF "anticipated" the shock.

We limit our focus to staff reports released in the period starting 6 years prior to the start of the shock and ending 5 years after the end of the shock episode. Within these reports, we focus on data on actual and forecasted production for reference years starting 3 years prior to the start of the shock and ending 3 years after the end of the shock episode. We are specifically extracting data on the forecasts that different countries had with respect to the commodities that they extract as a means to study how forecasts evolved over time. We are assuming that countries associate an anticipated decrease in production volume with an anticipated decrease in extraction revenue and furthermore that the Article IV consultative reports are mutually agreed upon by the authorities.

It is of note that article IV consultative reports only began in 1978, pursuant to the Second Amendment of the Articles of Agreement, which took effect on April 1st, 1978.

Thus, no articles are available for any country prior to that date. In fact, of the missing 7 out of 35 countries, 6 had their episodes start before 1978 and 1 joined the IMF after it experienced its decline episodes and thus no consultative report is available.

Along those lines, the definition of lack of anticipation, which applies to 31 out of the 35 episodes, is broadly understood as encompassing various case scenarios: reports where the commodity is not mentioned, which is usually for non-important commodities (6), no reports existing either due to it being before 1978 or due to country joining after the episode happened (7), no outlook mentioned (4), no anticipation at all, be it quantitative or qualitative (8). an imperfect anticipation where future production is set at t as opposed to t-1 (1), and cases of later anticipation where production expectations are adjusted only after the initial shock (5). Out of the the cases of both anticipation and lack of anticipation, the IMF forecast could take a qualitative or quantitative form, with the former consisting of either a mention of expected decrease/increase or a solid year-by-year binary coding of increases and decreases.

The reports can be found in two IMF databases, one of which is named eDOCS DM, and the other the IMF's internal search function iSearch. The first source, eDOCS, is typically used for older reports, and the second, iSearch, is typically used to access newer reports.

For each episode, in eDOCS, we input the relevant country's ISO-3 code in the "Country" field in the IMF Topic Search window. We then enter the starting and ending issue dates according to the period specified above. This query will return a list of search results, including information on the document title, series ID, issue date, and country, as well as a link to download the report as a PDF. We download the list as a PDF for record-keeping purposes, then download the staff reports as individual PDF files, usually a maximum of one per year, and ignoring corrigenda and supplementary files.

In the iSearch system, the approach is similar. We enter "Article IV" as the primary search term, then filter by country (including all names for the same country if there are multiple variations), year range as specified above, and file format as PDF. We export the results list as an Excel file for record-keeping purposes, then download the staff reports as individual PDF files, usually a maximum of one per year, and ignoring corrigenda and supplementary files.

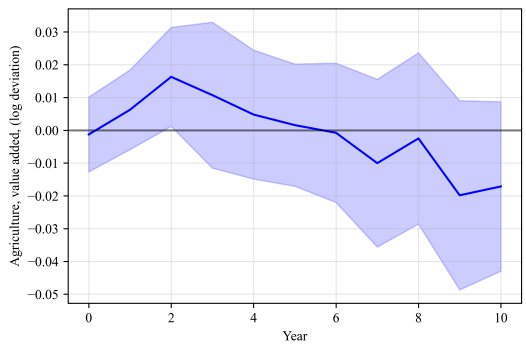
In the Excel file containing information on the shock episodes, we create a new sheet

for each shock episode. The sheet name nomenclature is the country's ISO-3 code, followed by the shock episode start year in YYYY format, then the commodity name. Report issue dates are columns, and data reference years are rows. The first column is a binary indicator which shows "1" if the reference year is included in the shock episode period, and "0" otherwise. Note that the "1" may be multiplied by a constant factor for chart-shading purposes. For instance, for a shock episode lasting from 2007 to 2011, the columns would include Article IV reports issued any time between 2001 and 2016 (inclusive), and the rows would include all data points referring to the years between 2004 and 2014 (inclusive).

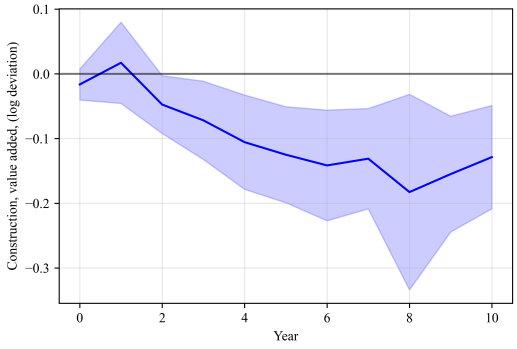
Data entry priority is as follows: (1) numerical values, first from tables then from text sections of the reports; (2) numerical values estimated from charts (extracted using open-source application WebPlotDigitizer); and finally in the absence of other information, (3) qualitative descriptions from the text (e.g., "decline" or "increase"). Page numbers are cited below the data; if only qualitative descriptions were found, more details will be given in the "Additional notes."

Note that, unless otherwise stated, years are assumed to be January–December, and references such as "last year" are always assumed to refer to the previous calendar year.

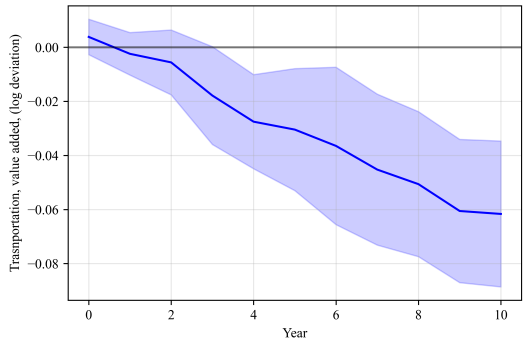
Industry Specific Analysis



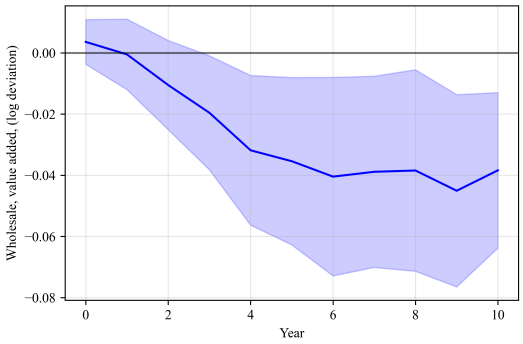
(a) Agriculture, value added



(b) Construction, value added



(c) Transportation, value added



(d) Wholesale, value added

Figure 3.23: Impulse response functions of real value added by sector to a shock episode.
Notes: Responses are generated using a shock episode normalized to an accumulated exogenous extraction decline of 40% over ten years. The shaded area reports 90 percent confidence intervals.

References

Abadie, A., Diamond, A., & Hainmueller, J. (2010). Synthetic control methods for comparative case studies: Estimating the effect of california's tobacco control program. *Journal of the American Statistical Association*, *105*(490), 493–505.

Abadie, A., & Gardeazabal, J. (2003). The economic costs of conflict: A case study of the basque country. *American Economic Review*, *93*(1), 113–132.

Acemoglu, D., Naidu, S., Restrepo, P., & Robinson, J. A. (2019). Democracy does cause growth. *Journal of Political Economy*, *127*(1), 47–100.

Acharya, S., & Dogra, K. (2020). Understanding hank: Insights from a prank. *Econometrica*, *88*, 1113–1158.

Allcott, H., & Keniston, D. (2018). Dutch Disease or agglomeration? the local economic effects of natural resource booms in modern america. *The Review of Economic Studies*, *85*(2), 695–731.

Almgren, M., Gallegos, J.-E., Kramer, J., & Lima, R. (2022). Monetary policy and liquidity constraints: Evidence from the euro area. *American Economic Journal: Macroeconomics*, *14*, 309–340.

Alvarez, L. J., Dhyne, E., Hoeberichts, M., Kwapil, C., Le Bihan, H., Lünnemann, P., Martins, F., Sabbatini, R., Stahl, H., Vermeulen, P., et al. (2006). Sticky prices in the euro area: A summary of new micro-evidence. *Journal of the European Economic association*, *4*(2-3), 575–584.

Arezki, R., Ramey, V., & Sheng, L. (2017). News shocks in open economies: Evidence from giant oil discoveries. *The Quarterly Journal of Economics*, *103*, 155.

Auclert, A. (2019). Monetary policy and the redistribution channel. *American Economic Review*, *109*, 2333–2367.

Auclert, A., Rognlie, M., Souchier, M., & Straub, L. (2021). Exchange rates and monetary policy with heterogeneous agents: Sizing up the real income channel. *NBER Working paper*.

Autor, D. H., & Dorn, D. (2013). The growth of low-skill service jobs and the polarization of the us labor market. *American economic review*, *103*(5), 1553–1597.

Bansal, R., Kiku, D., & Yaron, A. (2016). Risks for the long run: Estimation with time aggregation. *Journal of Monetary Economics*, *82*, 52–69.

Barro, R. J., & Sala-i-Martin, X. (1992). Convergence. *Journal of political Economy*, *100*(2), 223–251.

Bartelme, D. G., Costinot, A., Donaldson, D., & Rodriguez-Clare, A. (2019). *The textbook case for industrial policy: Theory meets data* (tech. rep.). National Bureau of Economic Research.

Bayer, C., Kriwoluzky, A., Müller, G. J., & Seyrich, F. (2024). A hank2 model of monetary unions. *Journal of Monetary Economics*, *147*, 103–179.

Becker, R., Gray, W., & Marvakov, J. (2021). Nber-ces manufacturing industry database (1958-2018, version 2021a).

Benmelech, E., & Bergman, N. K. (2009). Collateral pricing. *Journal of financial Economics*, *91*(3), 339–360.

Bernanke, B. S., Gertler, M., & Gilchrist, S. (1999). The financial accelerator in a quantitative business cycle framework. *Handbook of macroeconomics*, *1*, 1341–1393.

Bianchi, J. (2011). Overborrowing and systemic externalities in the business cycle. *American Economic Review*, *101*(7), 3400–3426.

Bianchi, J., & Mendoza, E. G. (2018). Optimal time-consistent macroprudential policy. *Journal of Political Economy*, *126*(2), 588–634.

Bilal, A. (2023). The geography of unemployment. *The Quarterly Journal of Economics*, *138*(3), 1507–1576.

Bilbiie, F. O. (2024). Monetary policy and heterogeneity: An analytical framework. *Review of Economic Studies*, rdae066.

Black, D., McKinnish, T., & Sanders, S. (2005). The economic impact of the coal boom and bust. *The Economic Journal*, *115*(503), 449–476.

Boppart, T. (2014). Structural change and the Kaldor facts in a growth model with relative price effects and non-Gorman preferences. *Econometrica*, *82*, 2167–2196.

Borusyak, K., Hull, P., & Jaravel, X. (2025). A practical guide to shift-share instruments. *Journal of Economic Perspectives*, *39*(1), 181–204.

BP. (2022). *Statistical review of world energy*. British Petroleum, London.

Brunnschweiler, C. N., & Bulte, E. H. (2008). The resource curse revisited and revised: A tale of paradoxes and red herrings. *Journal of Environmental Economics and Management*, *55*(3), 248–264.

Caliendo, L., Parro, F., Rossi-Hansberg, E., & Sarte, P.-D. (2018). The impact of regional and sectoral productivity changes on the us economy. *The Review of economic studies*, *85*(4), 2042–2096.

Caplin, A., Freeman, C., & Tracy, J. (1993). Collateral damage: How refinancing constraints exacerbate regional recessions.

Caselli, F. [Francesca], Grigoli, F., & Sandri, D. (2022). Protecting lives and livelihoods with early and tight lockdowns. *The BE Journal of Macroeconomics*, *22*(1), 241–268.

Caselli, F. [Francesco], & Coleman II, W. J. (2001). The us structural transformation and regional convergence: A reinterpretation. *Journal of political Economy*, *109*(3), 584–616.

Caselli, F. [Francesco], Koren, M., Lisicky, M., & Tenreyro, S. (2020). Diversification through trade. *The Quarterly Journal of Economics*, *135*(1), 449–502.

Cavalcanti, T., Da Mata, D., & Toscani, F. (2019). Winning the oil lottery: The impact of natural resource extraction on growth. *Journal of Economic Growth*, *24*(1), 79–115.

Center for Systemic Peace. (2022). *The polity project*. Center for Systemic Peace, Vienna, VA.

Chaney, T., Sraer, D., & Thesmar, D. (2012). The collateral channel: How real estate shocks affect corporate investment. *American Economic Review*, *102*(6), 2381–2409.

Chenery, H. B. (1960). Patterns of industrial growth. *The American Economic Review*, *50*(4), 624–654.

Clark, C. (1940). *The conditions of economic progress*. Macmillan.

Costinot, A., & Donaldson, D. (2012). Ricardo's theory of comparative advantage: Old idea, new evidence. *American Economic Review*, *102*(3), 453–458.

Costinot, A., Donaldson, D., Vogel, J., & Werning, I. (2015). Comparative advantage and optimal trade policy. *The Quarterly Journal of Economics*, *130*(2), 659–702.

Cravino, J., Lan, T., & Levchenko, A. A. (2020). Price stickiness along the income distribution and the effects of monetary policy. *Journal of Monetary Economics*, *110*, 19–32.

Cust, J., Harding, T., & Vezina, P.-L. (2019). Dutch Disease resistance: Evidence from indonesian firms. *Journal of the Association of Environmental and Resource Economists*, *6*(6), 1205–1237.

de Ferra, S., Mitman, K., & Romei, F. (2020). Household heterogeneity and the transmission of foreign shocks. *Journal of International Economics*, *124*, 103–303.

de Ferra, S., Mitman, K., & Romei, F. (2021). Why does capital flow from equal to unequal countries? *CEPR Discussion Paper*.

Deaton, A., & Muellbauer, J. (1980). *Economics and consumer behavior*.

Debortoli, D., & Galí, J. (2024). Heterogeneity and aggregate fluctuations: Insights from tank models. *National Bureau of Economic Research Working Paper*.

Desmet, K., & Rossi-Hansberg, E. (2014). Spatial development. *American Economic Review*, *104*(4), 1211–1243.

Do, Q.-T., & Levchenko, A. A. (2007). Comparative advantage, demand for external finance, and financial development. *Journal of Financial Economics*, *86*(3), 796–834.

Dorn, D. (2009). *Essays on inequality, spatial interaction, and the demand for skills* [Doctoral dissertation, Verlag nicht ermittelbar].

Dorrucci, E., Fidora, M., Gartner, C., & Zumer, T. (2022). The european exchange rate mechanism (erm ii) as a preparatory phase on the path towards euro adoption. *ECB Economic Bulletin*, *8*.

Eckert, F., Fort, T. C., Schott, P. K., & Yang, N. J. (2020). *Imputing missing values in the us census bureau's county business patterns* (tech. rep.). National Bureau of Economic Research.

Eckert, F., & Peters, M. (2022). *Spatial structural change* (tech. rep.). National Bureau of Economic Research.

Fajgelbaum, P., Morales, E., Suuarez Serrato, J. C., & Zidar, O. (2019). State taxes and spatial misallocation. *The Review of Economic Studies*, *86*(1), 333–376.

Federal Institute for Geosciences and Natural Resources. (2012). *Production database*.

Feenstra, R. C., Inklaar, R., & Timmer, M. P. (2015). The next generation of the penn world table. *American Economic Review*, *105*(10), 3150–82.

Ferrante, F., & Gornemann, N. (2022). Devaluations, deposit dollarization, and household heterogeneity. *International Finance Discussion Paper*.

Ferriere, A., Grübener, P., & Sachs, D. (2024). Optimal redistribution: Rising inequality vs. rising living standards. *CESifo Working paper*.

Feyrer, J., Mansur, E. T., & Sacerdote, B. (2017). Geographic dispersion of economic shocks: Evidence from the fracking revolution. *American Economic Review*, *107*(4), 1313–34.

Frankel, J. A. (2012). The natural resource curse: A survey of diagnoses and some prescriptions. *Commodity price volatility and inclusive growth in low-income countries*, 7–34.

Gan, J. (2007). Collateral, debt capacity, and corporate investment: Evidence from a natural experiment. *Journal of Financial Economics*, *85*(3), 709–734.

Ganong, P., & Shoag, D. (2017). Why has regional income convergence in the us declined? *Journal of Urban Economics*, *102*, 76–90.

Gautier, E., Conflitti, C., Faber, R. P., Fabo, B., Fadejeva, L., Jouvanceau, V., Menz, J.-O., Messner, T., Petroulas, P., Roldan-Blanco, P., et al. (2024). New facts on consumer price rigidity in the euro area. *American Economic Journal: Macroeconomics*, *16*, 386–431.

Gertler, M., & Karadi, P. (2011). A model of unconventional monetary policy. *Journal of monetary Economics*, *58*(1), 17–34.

Glaeser, E. L., Kallal, H. D., Scheinkman, J. A., & Shleifer, A. (1992). Growth in cities. *Journal of political economy*, *100*(6), 1126–1152.

Gornemann, N., Kuester, K., & Nakajima, M. (2021). Doves for the rich, hawks for the poor? distributional consequences of systematic monetary policy. *ECONtribute Discussion Paper*.

Graham, J. R., Leary, M. T., & Roberts, M. R. (2015). A century of capital structure: The leveraging of corporate america. *Journal of financial economics*, *118*(3), 658–683.

Groth, C., & Khan, H. (2010). Investment adjustment costs: An empirical assessment. *Journal of Money, Credit and Banking*, *42*(8), 1469–1494.

Guntin, R., Ottonello, P., & Perez, D. (2023). The micro anatomy of macro consumption adjustments. *American Economic Review*, *113*, 2201–2231.

Guo, X., Ottonello, P., & Perez, D. (2023). Monetary policy and redistribution in open economies. *Journal of Political Economy Macroeconomics*, *1*, 191–241.

Hall, R. E. (2004). Measuring factor adjustment costs. *The Quarterly Journal of Economics*, *119*(3), 899–927.

Hanson, G. H. (2023). Local labor market impacts of the energy transition: Prospects and policies. *NBER Working Paper*, *30871*.

Harding, T., Stefanski, R., & Toews, G. (2020). Boom goes the price: Giant resource discoveries and real exchange rate appreciation. *The Economic Journal*, *130*(630), 1715–1728.

Heathcote, J., Perri, F., & Violante, G. L. (2010). Unequal we stand: An empirical analysis of economic inequality in the united states, 1967–2006. *Review of Economic dynamics*, *13*(1), 15–51.

Heathcote, J., Perri, F., & Violante, G. L. (2020). The rise of us earnings inequality: Does the cycle drive the trend? *Review of economic dynamics*, *37*, S181–S204.

Heathcote, J., Perri, F., Violante, G. L., & Zhang, L. (2023). More unequal we stand? inequality dynamics in the united states, 1967–2021. *Review of Economic Dynamics*, *50*, 235–266.

Heblich, S., Nagy, D. K., Trew, A., & Zylberberg, Y. (2025). *The death and life of great british cities* (tech. rep.). National Bureau of Economic Research.

Herrendorf, B., Rogerson, R., & Valentinyi, Á. (2014). Growth and structural transformation. In P. Aghion & S. N. Durlauf (Eds.), *Handbook of economic growth* (pp. 855–941, Vol. 2). Elsevier.

Hong, S. (2020). *Emerging market business cycles with heterogeneous agents* (tech. rep.).

Imbs, J., & Wacziarg, R. (2003). Stages of diversification. *American economic review*, *93*(1), 63–86.

International Energy Agency. (2022). *World energy outlook* (International Energy Agency. Paris, France).

Jacobs, J. (1969). Strategies for helping cities. *The American Economic Review*, *59*(4), 652–656.

Jacobs, J. (1985). *Cities and the wealth of nations: Principles of economic life*. Vintage.

Jacobsen, G. D., & Parker, D. P. (2016). The economic aftermath of resource booms: Evidence from boomtowns in the american west. *The Economic Journal*, *126*(593), 1092–1128.

Jaimovich, N., Rebelo, S., & Wong, A. (2019). Trading down and the business cycle. *Journal of Monetary Economics*, *102*, 96–121.

Jaravel, X. (2018). The unequal gains from product innovations: Evidence from the u.s. retail sector. *Quarterly Journal of Economics*, *134*, 715–783.

Jarociński, M., & Karadi, P. (2020). Deconstructing monetary policy surprises—the role of information shocks. *American Economic Journal: Macroeconomics*, *12*, 1–43.

Jorda, O. (2005). Estimation and inference of impulse responses by local projections. *American Economic Review*, *95*, 161–182.

Kaplan, G., Moll, B., & Violante, G. L. (2018). Monetary policy according to hank. *American Economic Review*, *108*, 697–743.

Kiyotaki, N., & Moore, J. (1997). Credit cycles. *Journal of political economy*, *105*(2), 211–248.

Kongsamut, P., Rebelo, S., & Xie, D. (2001). Beyond balanced growth. *The Review of Economic Studies*, *68*(4), 869–882.

Korinek, A. (2011). The new economics of prudential capital controls: A research agenda. *IMF Economic Review*, *59*(3), 523–561.

Kose, M. A., Sugawara, N., & Terrones, M. E. (2020). Global recessions. *Policy Research Working Paper*, (9172).

Krueger, D., Perri, F., Pistaferri, L., & Violante, G. L. (2010). Cross-sectional facts for macroeconomists. *Review of Economic Dynamics*, *13*, 1–14.

Krugman, P. (1987). The narrow moving band, the Dutch Disease, and the competitive consequences of mrs. thatcher: Notes on trade in the presence of dynamic scale economies. *Journal of development Economics*, *27*(1-2), 41–55.

Kuznets, S. (1957). Quantitative aspects of the economic growth of nations: Ii. industrial distribution of national product and labor force. *Economic Development and Cultural Change*, *5*(4, Supplement), 1–111.

Lane, P. R., & Tornell, A. (1996). Power, growth, and the voracity effect. *Journal of Economic Growth*, *1*(2), 213–241.

Leite, C. A., & Weidmann, J. (1999). Does mother nature corrupt? natural resources, corruption, and economic growth. *IMF Working Paper*, *1999*(085).

Lustig, H., & Van Nieuwerburgh, S. (2010). How much does household collateral constrain regional risk sharing? *Review of Economic Dynamics*, *13*(2), 265–294.

Mauro, P. (1995). Corruption and growth. *The Quarterly Journal of Economics*, *110*(3), 681–712.

Mendoza, E. G. (2010). Sudden stops, financial crises, and leverage. *American Economic Review*, *100*(5), 1941–1966.

Mirzoev, M. T. N., Zhu, L., Yang, Y., Zhang, M. T., Roos, M. E., Pescatori, M. A., & Matsumoto, M. A. (2020). *The future of oil and fiscal sustainability in the gcc region*. International Monetary Fund.

Mitchell, B. R. (2013). *International historical statistics 1750-2005*. Springer.

Moffitt, R., & Zhang, S. (2018). Income volatility and the psid: Past research and new results. *AEA Papers and Proceedings*, *108*, 277–280.

Montiel Olea, J. L., & Plagborg-Møller, M. (2021). Local projection inference is simpler and more robust than you think. *Econometrica*, *89*, 1789–1823.

Moretti, E. (2010). Local multipliers. *American Economic Review*, *100*(2), 373–377.

Moroney, J. R., & Walker, J. M. (1966). A regional test of the heckscher-ohlin hypothesis. *Journal of Political Economy*, *74*(6), 573–586.

Morris-Levenson, J. (2022). *The origins of regional specialization* [Doctoral dissertation, The University of Chicago].

Olivi, A., Sterk, V., & Xhani, D. (2024). Optimal monetary policy during a cost of living crisis. *mimeo UCL*.

Oskolkov, A. (2023). Exchange rate policy and heterogeneity in small open economies. *Journal of International Economics*, *142*, 103750.

Peek, J., & Rosengren, E. S. (2000). Collateral damage: Effects of the japanese bank crisis on real activity in the united states. *American economic review*, *91*(1), 30–45.

Ramey, V. A. (2011). Identifying government spending shocks: It's all in the timing. *The Quarterly Journal of Economics*, *126*(1), 1–50.

Ramey, V. A., & Zubairy, S. (2018). Government spending multipliers in good times and in bad: Evidence from us historical data. *Journal of political economy*, *126*, 850–901.

Romer, C. D., & Romer, D. H. (2010). The macroeconomic effects of tax changes: Estimates based on a new measure of fiscal shocks. *American Economic Review*, *100*(3), 763–801.

Sachs, J. D., & Warner, A. M. (2001). The curse of natural resources. *European Economic Review*, *45*(4-6), 827–838.

Sala-i-Martin, X., & Subramanian, A. (2008). Addressing the natural resource curse: An illustration from nigeria. In *Economic policy options for a prosperous nigeria* (pp. 61–92). Springer.

Schmitz, C. J. (1979). *World non-ferrous metal production and prices: 1700-1976*. Cass.

Schwerhoff, G., & Stuermer, M. (2019). Non-renewable resources, extraction technology and endogenous growth. *Federal Reserve Bank of Dallas, Working Papers*, *2015*(1506).

Stiglitz, J. E., & Newbery, D. M. (1984). Pareto inferior trade.

Stock, J. H., & Watson, M. W. (2018). Identification and estimation of dynamic causal effects in macroeconomics using external instruments. *The Economic Journal*, *128*(610), 917–948.

Straub, L. (2019). Consumption, savings, and the distribution of permanent income. *Unpublished manuscript, Harvard University*.

Tornell, A., & Lane, P. R. (1999). The voracity effect. *American Economic Review*, *89*(1), 22–46.

United Nations. (2022). *Statistical database*. United Nations, New York.

U.S. Geological Survey. (2012). *Historical statistics for mineral and material commodities in the united states.*

Van der Ploeg, F., & Rezai, A. (2020). Stranded assets in the transition to a carbon-free economy. *Annual Review of Resource Economics*, *12*, 281–298.

Van der Ploeg, F., & Venables, A. J. (2012). Natural resource wealth: The challenge of managing a windfall. *Annual Review of Economics*, *4*(1), 315–337.

Walsh, C. (2023). The startup multiplier.

Watson, B., Lange, I., & Linn, J. (2023). Coal demand, market forces, and us coal mine closures. *Economic Inquiry*, *61*(1), 35–57.

World Bank. (2022). *New World Bank country classifications by income level: 2022-2023*. World Bank, Washington, DC.